Autumn Vol. 74 No. 3 (2023)

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 Professor Mark Flear. For the contribution of commentaries and notes, please contact Dr David Capper. For book reviews, contact Dr Clayton Ó Néill.

SUBSCRIPTIONS

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

NORTHERN IRELAND
LEGAL QUARTERLY

Autumn Vol. 74 No. 3 (2023)

Contents

Articles

Commentaries and Notes

Northern Ireland
Legal Quarterly
Autumn Vol. 74.3 (2023) 417–438
Articles DOI: 10.53386/nilq.v74i3.1065

What level of respect does opposition to same-sex marriage deserve in a democratic society?

Kenneth McK Norrie

University of Strathclyde
Correspondence email: kenneth.norrie@strath.ac.uk

ABSTRACT

The opening of marriage to same-sex couples shifted the legal debate from whether that should be done to the extent to which individuals and organisations can lawfully refuse to engage with same-sex marriage or can continue to argue against such marriages. This raises the question of the level of respect that needs to be shown to anti-LGBT (lesbian, gay, bisexual, transgender) views in a democratic society, a question that ought to have been, but was not, the central issue in *Lee v Ashers Baking Company Ltd*. By focusing on the right not to express a view that the defendants in that case did not hold, the United Kingdom(UK) Supreme Court avoided examining the views that they did, in fact, hold. The defendants' belief that LGBT people are sinful, manifested in their opposition to same-sex marriage, is in essence a belief in heterosexual superiority, which is a form of homophobia and therefore inconsistent with the values underlying the European Convention on Human Rights (ECHR), especially that of dignity. This article explores the level of respect such beliefs and their expression can expect to receive in the UK. It concludes that it should be no higher than 'toleration'. It will identify as the central flaw in the Supreme Court's approach that it afforded a higher level of respect than toleration, that it allowed the belief in heterosexual superiority to exempt the defendants from a legal obligation that would have to be met by those whose views on homosexuality were more in line with the values of the ECHR.

Keywords: LGBT rights; freedom of belief; freedom of expression; discrimination; dignity; homophobia; proportionality; same-sex marriage.

INTRODUCTION

The opening of marriage to same-sex couples throughout the United Kingdom (UK)[1] resolved the headline question, but it did not end the debate, which has shifted to the extent to which opponents can

1 Marriage (Same Sex Couples) Act 2013 (for England & Wales); Marriage and Civil Partnership (Scotland) Act 2014 (for Scotland); Marriage (Same-Sex Couples) and Civil Partnership (Opposite-Sex Couples) (Northern Ireland) Regulations 2019 (SI 2019/1514) and the Marriage and Civil Partnership (Northern Ireland) Regulations 2020 (SI 2020/742).

either refuse to engage with any involvement in such marriages or can legitimately continue to speak out against such marriages. A recent manifestation of this continuing debate is found in *303 Creative LLC et al v Elenis*,[2] where in June 2023 the United States (US) Supreme Court decided that a web-designer for wedding planners would be entitled to limit the services she offered to those planning opposite-sex weddings only, for otherwise she would be forced to express a belief of the state (that marriage could legitimately be entered into by a same-sex couple) rather than one she possessed herself (that such a marriage was contrary to her god's will). At the oral hearing in December 2022, Justice Alito (who subsequently joined the majority opinion, penned by Justice Gorsuch) had sought to distinguish between 'decent and respectable' opposition to same-sex marriage and (presumably unacceptable) opposition to interracial marriage.[3] Also in December 2022 President Biden signed into law legislation to ensure that same-sex marriages continue to be recognised throughout the US, but which contains a provision allowing faith groups to refuse goods, services and accommodations in connection with same-sex marriages.[4]

It is timely, therefore, to revisit the implications for this ongoing debate of the UK Supreme Court's decision in *Lee v Ashers Baking Company Ltd*.[5] This was the case in which Mr Gareth Lee had claimed discrimination when a Belfast retail baking company refused to supply him with a cake bearing the words 'Support Gay Marriage' in its icing. The claim based on sexual orientation discrimination need not detain us here: it was dismissed because, in the words of Lady Hale, the baking company's objection 'was to the message and not to any particular person or persons'.[6] It is the claim for discrimination on the ground of political opinion, also dismissed by the Supreme Court, that is the focus of this article.

The message on the cake was clearly the manifestation of a political view, which the customer sought to express at a time when the debate on opening marriage to same-sex couples was very much

2 *303 Creative LLC v Elenis* 600 US 570 (2023).

3 See https://www.oyez.org/cases/2022/21-476 (last accessed 19 November 2023).

4 Respect for Marriage Act 2022, Public Law 117-228, s 6. See also, to similar but more limited effect, s 16 of the Marriage and Civil Partnership (Scotland) Act 2014.

5 [2018] UKSC 49.

6 Ibid [34].

live, and contentious, in Northern Ireland.[7] Mr Lee argued that the refusal to accept his order amounted to discriminatory treatment on the basis of his political opinion, contrary to the Fair Employment and Treatment (Northern Ireland) Order 1998.[8] However, instead of exploring the level of protection that Mr Lee's political opinion was entitled to, the Supreme Court turned the spotlight away from the customer and onto the baking company's (and its owners') beliefs. It then resolved the case on the basis of the baking company's right not to be forced to express views that it did not hold, a sleight of hand that relieved the Supreme Court of the task of scrutinising the level of protection to be afforded to the actual beliefs held by the baking company and its owners. In January 2022 the Fourth Section of the European Court of Human Rights rejected as manifestly unfounded an application by Mr Lee claiming that his rights under articles 8, 9 10 and 14 of the European Convention on Human Rights (ECHR) had been infringed, because he had not exhausted his domestic remedies in respect of his own Convention rights. The fact that Convention rights had been fully explored by the Supreme Court was dismissed as irrelevant because the domestic courts had been asked to balance Mr Lee's rights under domestic law with the defendants' Convention rights, and he could not now ask the European Court to balance his own Convention rights with those of the defendants.[9] This allowed the European Court to avoid, like the UK Supreme Court, scrutinising the baking company's actual beliefs the manifestation of which Mr Lee was complaining about.

It is that scrutiny that I seek to undertake here. Both the Supreme Court and the European Court assumed that the baking company's belief – that same-sex marriage should not be supported because it was wrong and sinful – was entitled to respect and protection under articles 9 and 10 of the ECHR. As we will see below, the Convention permits substantial limitations on manifestations of racist (or anti-Semitic or Islamophobic) beliefs and the belief, for example, that mixed-race marriages should not be permitted would be afforded little if any respect in a decent and democratic society. The purpose of this article is to interrogate whether opposition to same-sex

7 The Northern Ireland regulations cited in n 1 above were made a year after the case, following the UK Parliament's imposition of a requirement that the Secretary of State for Northern Ireland (a UK Government minister) do so: Northern Ireland (Executive Formation etc) Act 2019, s 8.

8 SI 1998/3162 (NI 21).

9 *Lee v United Kingdom* Appl 18860/19, 6 January 2022, [75].

marriage, when a manifestation of homophobia,[10] falls into the same category as racism and the extent to which the assertion that marriage should be limited to opposite-sex couples remains within the realms of acceptable – or respect-worthy – political debate entitled to protection under the ECHR.

THE IMPORTANCE OF DIGNITY AND TOLERANCE

The level of respect that is to be afforded to any manifestation of belief, or freedom of expression, is determined not only by the words of the substantive provisions of the ECHR but also by its spirit and underlying values. Beliefs 'deserve little, if any, protection if their content is at odds with the democratic values of the Convention system'.[11] These democratic values include 'pluralism', that is to say the 'harmonious interaction of persons and groups with varied identities',[12] 'tolerance, social peace and non-discrimination'[13] and, perhaps most significant for the buttressing of lesbian, gay, bisexual and transgender (LGBT) rights, dignity,[14] which is increasingly recognised as inherent in many of the substantive articles of the ECHR.[15] In *Bouyid v Belgium*, the European Court pointed out that, although the ECHR itself does not mention the concept of dignity,[16] 'the Court has emphasised that respect for human dignity forms part of the very essence of the Convention,

10 It is important to acknowledge that not all opposition to same-sex marriage is necessarily homophobic: see, for example, Kenneth Norrie, 'Marriage is for heterosexuals: may the rest of us be saved from it' (2000) 12 Child and Family Law Quarterly 363, who argues that some essential differences between same-sex and opposite-sex couples are ignored by applying one institution (marriage) to both; and Nancy Polikoff, 'We will get what we ask for: why legalizing gay and lesbian marriage will not dismantle the legal structure of gender in every marriage' (1993) 79 Virginia Law Review 1535 and *Beyond (Straight and Gay) Marriage: Valuing All Families under the Law* (Beacon Press 2008), who argues in the former piece that marriage is irredeemably gendered and in the latter piece suggests that the focus on marriage is not the best means to ensure genuine and universal LGBT equality throughout society and particularly in family law. Neither author (a gay man and a lesbian) is arguing for LGBT people to be treated less well by the law than non-LGBT people.

11 *Pastörs v Germany* Appl 55225/14, 3 October 2019, [47].

12 *Baczkowski v Poland* (2009) 48 EHRR 19, [62], [63].

13 *Norwood v United Kingdom* (2005) 40 EHRR SE11.

14 See Peter Laverack, 'The indignity of exclusion: LGBT rights, human dignity and the living tree of human rights' [2019] EHRLR 172.

15 See Sebastian Heselhaus and Ralph Hemsley, 'Human dignity and the European Convention on Human Rights' in Paolo Becchi and Klaus Mathis (eds), *Handbook of Human Dignity in Europe* (Springer 2019) 969–992.

16 Though it does appear in the preamble to optional Protocol 13.

alongside human freedom'; it follows that 'any interference with human dignity strikes at the very essence of the Convention'.[17]

Tolerance of others, their actions, opinions and beliefs – and especially of their immutable characteristics such as race and sexual orientation – is the ultimate guarantor of respect for the dignity of others and thereby the guarantor of true equality. This means that philosophies and beliefs (irrespective of their source) that encourage intolerance of others, that denigrate different identities by denying or diminishing the dignity of others, that seek conflict and division, are contrary to the values underlying the ECHR and are worthy of little or no respect in a democratic society. But it does not mean that they are entitled to none.

DIFFERING LEVELS OF RESPECT

It is settled law that, while the right guaranteed by article 9 of the ECHR to freedom of thought, conscience and religion is unqualified, the right to manifest thought, conscience or religion and the right of expression, including the right guaranteed by article 10 to receive and impart information and ideas without interference, may be subject to limitations. Indeed, article 17 of the Convention itself provides that certain extreme views may be deprived completely of any protection, because they are not worthy, in a democratic society, of any respect whatsoever.[18]

Article 17 has a radical effect, and for that reason its application requires a high threshold: it removes entirely the protections offered by the ECHR and so may be invoked only when there is an attempt to use Convention rights to challenge the very notion of democracy the protection of which is the Convention's overarching purpose.[19] In the context of freedom of speech, article 17 is normally used to prevent those seeking to propagate the worst forms of hate speech, or the transmission of ideas that encourage violence, hatred or social unrest, from claiming that article 10 gives them the right to do so.[20] The

17 (2016) 62 EHRR 32 [89], [101].

18 Art 17: 'Nothing in this Convention may be interpreted as implying, for any State, group or person any right to engage in any activity or perform any act aimed at the destruction of any of the rights or freedoms set forth herein or at their limitation to a greater extent than is provided for in the Convention.'

19 *Vajnai v Hungary* (2010) 50 EHRR 44, [21]–[26]; *Ibragimov v Russia* Appl 1413/08, 28 August 2018, [62].

20 In *Reference by Attorney General of Northern Ireland – Abortion Services (Safe Access Zones) (Northern Ireland) Bill* [2022] UKSC 32, [54], the UK Supreme Court pointed out that conduct will fall outwith the protection of the Convention by operation of art 17 if it 'involves violent intentions, or incites violence, or otherwise rejects the foundations of a democratic society'.

typical example is Holocaust denial, or support for Nazi ideology.[21] In *Norwood v United Kingdom*[22] a member of the British National Party, a far-right group with strongly racist beliefs, had been convicted of an offence under section 5 of the Public Order Act 1986 having displayed in his window a poster showing a photograph of the New York Twin Towers aflame on 9/11, and the words 'Islam out of Britain'. When he complained to the European Court that his conviction was an infringement of his article 10 right to freedom of expression the Court dismissed the application as manifestly unfounded, concluding that his actions were 'incompatible with the values proclaimed and guaranteed by the Convention, notably tolerance, social peace and non-discrimination'.[23]

However, the Convention as a whole does not operate as a binary, with expressions of views either respected fully by articles 9 and 10 or rejected entirely by article 17. The respect that requires to be shown to views is, rather, determined by a sliding scale with differing levels of respect being required in different contexts. States are allowed to impose limitations on the expression of views that would not engage article 17, and the validity of these limitations is determined by the familiar proportionality analysis. Individuals may be entitled to hold homophobic views, but the respect that states must show to these views will differ, depending upon a variety of factors.

In, for example, *Lilliendahl v Iceland*[24] a man had been convicted of hate speech against LGBT people, and though the threshold for article 17 was held not to have been reached his comments were found to 'promote intolerance and detestation of homosexual persons' and so legitimately subject to restrictions. His complaint under article 10 was rejected as manifestly ill-founded, not because of article 17 but because article 10 itself contains another, and more easily satisfied, source of the requirement that the right to freedom of expression must be exercised consistently with the underlying values of the European Convention. Article 10(2) provides: 'the exercise of the freedom of expression carries

21 The first unequivocal recognition that Holocaust denial would be excluded completely from ECHR protection by art 17 is in *Lehideux v France* (2000) 30 EHRR 665, [47] (though in that case itself the contentious attempts to rehabilitate the reputation of the leader of Vichy France, Pétain, were held not to amount to Holocaust denial). See generally Paolo Lobba, 'Holocaust denial before the European Court of Human Rights: evolution of an exceptional regime' (2015) 26 European Journal of International Law 237.

22 See *Norwood* (n 13 above).

23 See also *Glimmerveen v Netherlands* (1982) 4 EHRR 260, [16] for a further description of the scope of art 17. That case involved the banning of a political party that sought to remove all non-white persons from the Netherlands.

24 Appl 29297/18, 12 May 2020.

with it duties and responsibilities',[25] and the applicant in *Lilliendahl* was held to have abused his responsibilities by speaking in the manner he did. Uniquely within the European Convention, article 10(2) explicitly imposes responsibilities on individuals (as well as the more familiar positive obligations on states that are implicit in many of the substantive articles of the Convention). In *Vejdeland v Sweden*[26] various individuals entered a secondary school and distributed leaflets describing homosexuality as a 'deviant sexual proclivity' and alleged that LGBT rights groups 'played down paedophilia'. Rejecting their complaint that their conviction for agitation against a national or ethnic group was a breach of their article 10 rights, the European Court emphasised the obligations on the applicants that arise alongside the rights in article 10, 'one such obligation being, as far as possible, to avoid statements that are unwarrantably offensive to others, constituting an assault on their rights'.[27] Though the concept of responsibility remains vague in the European Court's jurisprudence, it would seem that the obligation on speakers is to be aware of the context in which they speak, and to moderate that speech when necessary to avoid compromising the underlying values of the Convention.[28] Speakers need, in other words, to be aware of the potential for their words to lead to unwarranted hurt in others, as well as to violence or social unrest: 'even paying due regard to the qualified right to freedom of expression, people cannot expect to be protected if their core belief involves violating others' dignity and/or creating an intimidating, hostile, degrading, humiliating or offensive environment for them'.[29]

While the right to hold a belief is absolute under article 9, that does not mean that the state, when adjudicating on its manifestation, or the expression of the belief under article 10, is prohibited from evaluating how close the belief itself is to the democratic values underpinning the Convention. The further the belief is from the core values of the

25 See Claire Moran, 'Responsibility and freedom of speech under article 10' [2020] European Human Rights Law Review 67, who warns against an over-reliance on art 17 and would prefer the focus to be on the limitations in art 10.

26 *Vejdeland v Sweden* (2014) 58 EHRR 15. See also *Hammond v DPP* [2004] EWHC 69 (Admin) where an evangelical Christian preacher, while preaching in a public street, displayed a sign 'Stop Homosexuality. Stop Lesbianism'. His conviction under s 5 of the Public Order Act 1986 was upheld and his defence based on arts 9 and 10 rejected as his actions were not a reasonable exercise of the freedoms protected by these articles.

27 *Vejdeland v Sweden* (n 26 above) [57].

28 A clear example of this context-dependency is the responsibility of members of the judiciary to show 'maximum discretion' in expressing views in public due to the 'special role in society of the judiciary, which, as a guarantor of justice, a fundamental value in a law-governed society, must enjoy public confidence if it is to be successful in carrying out its duties': *Baka v Hungary* (2017) 64 EHRR 6, [164].

29 *Forstater v CGD Europe* [2019] 12 WLUK 516, [87] (Employment Tribunal).

Convention, the easier it will be for the state to establish a justification for any restriction on its manifestation or expression. In *Pastörs v Germany*[30] a German *Land* MP complained of having been convicted of Holocaust denial[31] as a result of various statements he had made in the *Land* Parliament. The European Court of Human Rights, while acknowledging that the existence of that offence in German criminal law amounted to a restriction on the right to freedom of expression, said this:

> The applicant sought to use his right to freedom of expression [article 10] with the aim of promoting ideas contrary to the text and spirit of the Convention. This weighs heavily in the assessment of the necessity of the interference … While interferences with the right to freedom of expression call for the closest scrutiny when they concern statements made by elected representatives in Parliament, utterances in such scenarios deserve little, if any, protection if their content is at odds with the democratic values of the Convention system.[32]

Likewise, the strength of the language used to express the view, the context in which it is expressed and its capacity to cause offence and social unrest will be central to the proportionality analysis. Actions or statements that are likely to offend significant numbers of people are manifestations of belief whose prohibition may well be proportionate to the legitimate aim of avoiding social conflict, as with the street preacher sermonising about the sinfulness of homosexuality in *Hammond v DPP*.[33] But since a tolerant society tolerates the offensive (and indeed the intolerant),[34] to lose ECHR protection the belief or expression thereof must do more than simply offend: it must have the potential to cause *harm*, such as 'expressions that seek to spread, incite or justify hatred based on intolerance, including religious intolerance'.[35] Though offensiveness alone is not sufficient to deny a belief or its expression of the protections in articles 9 and 10, the Convention is designed to do far more than discourage social unrest, for the parameters of 'harm' here are wider than violence or civil disturbance. It will normally be enough to justify restrictions if the rights, broadly conceived, of other individuals would be detrimentally affected without the prohibition. In *Vejdeland v Sweden*,[36] mentioned above, the European Court found

30 Appl 55225/14, 3 October 2019.

31 A crime in most countries in Europe (though not, specifically, in the UK) and in Israel.

32 Appl 55225/14, 3 October 2019, [46]–[47]. See also [37] for the relevance of the concepts in art 17 in the proportionality analysis.

33 [2004] EWHC 69 (Admin).

34 As Lord Walker put it (quoting counsel) in *R (Williamson) v Secretary of State for Education* [2005] UKHL 15, [60]: 'In matters of human rights the court should not show liberal tolerance only to tolerant liberals.'

35 *ES v Austria* (2019) 69 EHRR 4, [43].

36 *Vejdeland* (n 26 above).

that, while the anti-LGBT leaflets the applicants had distributed in a high school did not directly recommend individuals to commit hateful acts, they did make serious and prejudicial allegations, and therefore the conviction of the applicants was a proportionate interference with article 10:

> The Court reiterates that inciting to hatred does not necessarily entail a call for an act of violence, or other criminal acts. Attacks on persons committed by insulting, holding up to ridicule or slandering specific groups of the population can be sufficient for the authorities to favour combating racist speech in the face of freedom of expression exercised in an irresponsible manner.[37]

Inherent in the Convention is the understanding that there is no inconsistency between sincerely holding to one belief while respecting the right of others to hold a contrary belief: indeed showing respect to the beliefs of others is the only way of ensuring the inherent dignity of every human person. It follows that the more intolerant of other views a belief is, the easier the court will find it to hold the belief inconsistent with the Convention aim of fostering tolerance and pluralism, and the more readily the limitations on its manifestation can be found justified.

That messages of tolerance will find it far easier to be considered consistent with the Convention and entitled to a higher degree of respect than messages of intolerance, and that restrictions on the latter are therefore more likely to be found to be proportionate than restrictions on the former, is well illustrated in *R (Core Issues Trust) v Transport for London*.[38] Transport for London (TfL) had displayed adverts on the sides of London buses, placed by Stonewall UK (an LGBT rights organisation) with the words 'Some People are Gay. Get Over it'. Shortly thereafter, TfL refused to display adverts on its buses placed by Core Issues Trust, an organisation that claims the mission to support gay people seeking to become non-gay people (by a process the organisation calls 'sexual re-orientation' and which is normally referred to as 'gay conversion therapy').[39] The organisation challenged, by way of judicial review, TfL's refusal of their advert arguing (*inter alia*) that the refusal interfered with their article 10 right to freedom of expression. The challenge was rejected by the Court of Appeal which saw a crucial difference between the two adverts. Lord Dyson MR explained that

37 Ibid [55].

38 *R (Core Issues Trust)* [2014] EWCA Civ 34.

39 The UK Government published plans to ban 'these coercive and abhorrent practices' in England and Wales: *Banning Conversion Therapy Consultation* 21 October 2021. Likewise, 'The Scottish Government has committed to introducing legislation by the end of 2023 which will end conversion practices as comprehensively as possible within devolved powers': 'Ending conversion practices' (*Gov.scot* 18 November 2021).

Stonewall's advert had been designed to discourage discrimination, while the Trust's advert would *encourage* discrimination:

> The restrictions are justified in view of the prominence of the advertisements and the fact that they would be seen by, and cause offence to, large numbers of the public in central London. Moreover, for those who are gay, the advertisements would be liable to interfere with the right to respect for their private life under article 8(1).
>
> ... I agree with the judge that the advertisement is liable to encourage homophobic views *and homophobia places gays at risk* ...
>
> I consider that the Stonewall advertisement was intended to promote tolerance of homosexuals and discourage homophobic bullying ... The Trust's advertisement was a riposte to the 'gay acceptance' message promoted by Stonewall and would have been seen (and was seen) as countering that message and encouraging 'gay rejection' by implying offensively and controversially that homosexuality can be cured.[40]

The importance of this judgment for our purposes is not only that it recognises that encouraging homophobic views is dangerous for LGBT people but also that, for that reason, limitations on free speech designed to prevent homophobia are likely to be proportionate to their protective aims – and are likely thereby to be consistent with the ECHR.

HOMOPHOBIA: AN ASSUMPTION OF SUPERIORITY

'Homophobia places gays at risk.' This is reason enough to show that its manifestation, whether by direct expression or by implication from acts or omissions, will be inimical to the values that underpin the ECHR. But it is wrong to restrict our understanding of homophobia only to speech or behaviour that indicates mindless hatred of, extreme prejudice against, or the encouraging of violence towards LGBT people. LGBT people are exposed to risk in many more insidious ways than by attacks on our physical safety. Public attitudes and cultural mind-sets can (and for many long centuries did) create a social atmosphere that is toxic for LGBT people, one with a substantial ability to impact destructively on our emotional and mental wellbeing,[41] our sense of self and our place in society. The word 'homophobic' is today widely understood to encompass beliefs and expressions that are less obviously destructive, and it is commonly used to describe even moderately expressed

40 *R (Core Issues Trust)* (n 38 above) [84]–[85], [88] (emphasis added).

41 See, for example, K Schreiber and H Hausenblas, 'Why are suicide rates higher among LGBTQ youth?' (*Psychology Today UK* 12 October 2017). The Mental Health Foundation refers to a survey by Stonewall that half of all LGBTIQ+ have experienced depression, that one in eight have attempted suicide and almost half of all trans people have considered suicide: 'LGBTIQ+ people: statistics'.

anti-LGBT viewpoints – anything indeed that suggests that gay and lesbian people are of less worth, due to their sexual orientation, than heterosexual people.[42] Homophobia, in essence, is the assumption of heterosexual superiority – the spoken or unspoken belief that the holder of the belief, because heterosexual, is morally better and socially more valuable than those who are not, irrespective of any intent to discriminate against or otherwise harm any individual or group. Lady Hale, in *Lee v Ashers Baking Company Ltd*, spoke to the lived experience of millions when she said that 'it is deeply humiliating, and an affront to human dignity, to deny someone a service because of that person's race, gender, disability, sexual orientation or any of the other protected personal characteristics'.[43] The affront to dignity lies in the necessary implication that a person – or a whole community – is worth less than others, merely on account of their race, gender, disability or sexual orientation. Homophobia in this broader sense, because it denies the dignity of those it assumes to be inferior, offends 'the very essence of the Convention'[44] just as much as violent or extreme homophobia does, if in a different (but more insidious) way: it causes real and enduring harm to every LGBT person. So even if the owners of the Ashers Baking Company Ltd act respectfully to LGBT individuals that they come across in their professional and personal lives, and even when it would not cross their minds (as good Northern Ireland Christians) to encourage violence or hatred against anyone, their belief in their own heterosexual superiority may justly be described as homophobic, and as harmful to those who do not share their sexuality. *Lee v Ashers Baking Company Ltd* may not have been about an act of discrimination against someone because of their sexual orientation, but it was fundamentally about the extent to which homophobic belief was worthy of respect in a democratic society. Though obscured by the Supreme Court's focus on the defendants' right not to express a belief that they did not actually hold, the defendants were in essence asserting their right to deny the equal worth of homosexuality and heterosexuality as aspects of the human condition.

This is no different from the claim to racial superiority that lies at the heart of racism, and courts often point to homophobia and racism

42 In *Campbell v Dugdale* [2020] CSIH 27 the Court of Session accepted as fair comment a description of a tweeter as 'homophobic' when he had made a clumsy joke in poor taste at the expense of a gay MP.

43 *Lee v Ashers* (n 5 above) [35].

44 *Pretty v United Kingdom* (2002) 35 EHRR 1, [65].

being regarded in the same light, and as equally unworthy of respect.[45] But *Lee v Ashers Baking Company Ltd* reveals that the analogy is not (yet) accepted as exact. While racism that falls short of violent hatred has long been regarded as incompatible with Convention values, the Supreme Court's approach creates doubt as to whether homophobia falling short of violent hatred is equally incompatible. Few would deny the racism in a remark like: 'I wish black people well, and no harm, but I would not want to associate with them myself because they are racially inferior to me', because that is a claim to superiority based on no factor remotely relevant to worth and is indifferent to the very real harm that such assertion of superiority does to the dignity of all black people. Replacing 'black' with 'LGBT', and 'racially' with 'sexually', should not make the remark any more acceptable in a democratic society, nor entitled to any greater respect or protection. But the Supreme Court has – unintentionally I accept – cast doubt on that proposition.

Imagine the following two scenarios:

Scenario one

> Ms Aristel Basquet goes to a bakery and asks for a cake to be decorated with the words 'Black Lives Matter', which she intends to share at a rally supporting that movement. The baker, very politely, declines on the ground that it is contrary to her own philosophical belief that, in fact, white lives matter more than black lives and that it is white people, more than black people, who today suffer true discrimination. Ms Basquet complains of discrimination in the provision of goods and services and in defence the baker pleads article 10 of the ECHR, asserting that it would breach her right of free speech to be forced to give expression to a belief that she does not, in fact, hold.

Scenario two

> Mr Pierre Makarel goes to a bakery and asks for a cake to be decorated with the words 'Islam Out of UK', which he intends to share at a rally of neo-Nazis. The baker, not very politely, declines on the ground that it is contrary to his own philosophical belief that racial and religious hatred should always be resisted. Mr Makarel complains of discrimination and in defence the baker

45 In *Grainger plc v Nicholson* [2010] 2 All ER 253, [28], Burton J offered as examples of beliefs that would not be worthy of respect in a democratic society, would not be compatible with human dignity and would conflict with the fundamental rights of others 'a racist or homophobic political philosophy'. The European Court in *Vejdeland v Sweden* (n 26 above) [55], a case about homophobic belief, gave 'racist speech' as an example of speech that would be an irresponsible exercise of freedom of expression.

> pleads article 10 of the ECHR, asserting that it would breach his right of free speech to be forced to give expression to a belief that he does not, in fact, hold.

Are the beliefs held by these bakers each entitled to the same level of respect under articles 9 and 10? A proportionality analysis would readily allow the courts to give different answers in these two scenarios. In the first, the belief held by the baker (white superiority), being inconsistent with the values of tolerance and pluralism that underpins the Convention, should not receive a level of respect that would allow it to be used as a defence to a claim for discrimination. In the second, the baker's belief (in the importance of resisting racial and religious hatred) is entirely consistent with both the text and the spirit of the Convention and for that reason is entitled to a higher degree of respect in a democratic society than the belief in the first scenario, *including its manifestation by use as a defence to a claim for discrimination*. If so, then the baker in the first scenario should be unable to rely on article 10 and so would be obliged to fulfil the order (or leave the business of offering goods and services to the public), while the baker in the second scenario would be able to use article 10 to protect himself from being forced to express a view with which he profoundly disagrees.[46]

Now, in *Lee* the Supreme Court failed to subject the actual belief of the owners of the Ashers Baking Company Ltd, or how they manifested their belief, to any assessment of its consistency with ECHR values, because it diverted its attention to the belief that the owners did *not* hold, and which they claimed they were being forced to express. Had the focus been on the actual belief of the owners, the Supreme Court might well have realised that their belief was analogous to the belief of the baker in the first scenario – analogous to racism, in other words, since both beliefs are founded on an (unmerited) assumption of superiority, notwithstanding that their refusal to supply the cake was not in itself an expression of hatred nor, directly, the promotion of violence or prejudice. Instead, the owners' belief (in their own heterosexual superiority) was treated, by default, as analogous to the

46 This assumes, for the sake of present argument, that the Supreme Court were right in their assumption that a supplier of goods decorated with words is expressing the content of the words. I previously challenged that assumption in 'Case and comment: *Lee v Ashers Baking Company Ltd*' (2019) Juridical Review 88, 92: 'The Royal Mail is not endorsing or even expressing the views of political parties who use its services to distribute their propaganda: it is simply providing a medium. Broadcasters are not themselves expressing a view when they give airtime to those who offer opinions. No-one really believes that a shop selling cards, or cakes, with "Happy Birthday" on them is itself expressing birthday greetings to anyone.'

baker's belief in the second scenario – it was a belief entitled to a level of respect that allowed the owners to use it to claim an exemption from the general law prohibiting discrimination. By doing so, the Supreme Court afforded homophobic belief that was not directly violent more respect than it would likely afford racism that is not directly violent, an approach inconsistent with the position of the European Court that discrimination based on sexual orientation is as serious as discrimination based on 'race, origin or colour'.[47]

THE LEVEL OF PROTECTION PROPERLY AFFORDED TO OPPOSITION TO SAME-SEX MARRIAGE

It is likely that few people who oppose same-sex marriage self-identify as 'homophobic'. But just as opposition to mixed-race marriage is unquestionably racist, as a proxy for the belief in racial supremacy, so too opposing same-sex marriage, when founded on the belief in heterosexual superiority,[48] is necessarily homophobic. Both oppositions are therefore deeply suspect as being inconsistent with the underlying values of the European Convention, in particular that of dignity. But that does not mean that expressing opposition to same-sex marriage is entitled to no protection at all under the Convention. I am far from arguing that opposing the opening of the institution of marriage to same-sex couples is as destructive of Convention values as Holocaust denial or support for neo-Nazism: it does not, in itself, activate article 17,[49] nor fail to satisfy the fifth *Grainger*

47 *Smith and Grady v United Kingdom* [2000] 29 EHRR 493, [97]. See also *Karner v Austria* (2004) 38 EHRR 24, [37]: 'Just like differences based on sex, differences based on sexual orientation require particularly serious reasons by way of justification.' In *Ratzenböck and Seydl v Austria* Appl 28475/12, 26 October 2017 [32], the court reiterated that 'differences based solely on sexual orientation are unacceptable under the Convention'. In *Campbell v Dugdale* 2019 SLT (Sh Ct) 141, [86], Sheriff Ross described, in the context of an action for defamation, an allegation of homophobia as 'toxic' and bearing comparison with 'racist' or 'holocaust denier'.

48 See n 10 above.

49 See *Lilliendahl v Iceland* Appl No 29297/18, 12 May 2020 where art 17 was held not to be activated by comments designed to promote detestation of homosexuals. This was a hate crime far more explicitly homophobic than opposition to same-sex marriage, and the European Court of Human Rights has, in fact, never used art 17 to deprive homophobic speech of any ECHR protection.

criterion.[50] Nor am I arguing that opposing same-sex marriage should be criminalised as a hate crime. Free debate is so deeply rooted in the needs of democracy that the courts are, rightly, very reluctant to condemn contributions to political debate as unworthy of any respect at all in a democratic society. In the well-known words of Sedley LJ in *Redmond-Bate v DPP*:[51] 'Free speech includes not only the inoffensive but also the irritating, the contentious, the eccentric, the heretical, the unwelcome and provocative, provided it does not tend to provoke violence. Freedom only to speak inoffensively is not worth having.' The right to offend is itself an essential element of democracy,[52] and that means that even homophobic beliefs and their expression must be entitled to some level of ECHR respect and protection. The question is, what level?

There are a number of cases which show that, as a minimum, speech involving claims of heterosexual superiority has to be tolerated, in the sense of being permitted to be expressed. In *Re Christian Institute's Application for Judicial Review*[53] the applicants (a group of Christian churches and organisations) sought the nullification of a Northern Irish statutory instrument[54] prohibiting discrimination and harassment on the ground of sexual orientation: they argued that granting this protection to the LGBT community would interfere with the applicants' conception of Christian beliefs (by requiring them and their fellow adherents to treat gay people with the same level of respect that they would show to non-gay people). The application was granted in part, and, though that was mainly due to flaws in the consultation process, Weatherup J took pains to point out:

---

50 In *Grainger plc v Nicholson* [2010] 2 All ER 253 Burton J set out the criteria that needed to be satisfied before an opinion qualified as a philosophical belief worthy of protection. The fifth criterion was that it be not incompatible with human dignity and not conflict with the fundamental rights of others. It is worth noting that Burton J, though he did not have opposition to same-sex marriage in mind explicitly suggested at [28] that 'a homophobic political philosophy' would indeed fail his fifth criterion. In *Forstater v CGD Europe* [2021] 6 WLUK 104 the Employment Appeal Tribunal said at [66]: 'even comments which are "serious, severely hurtful and prejudicial", or which promote intolerance and detestation of homosexuals would not fall outside the scope of art 10 altogether. However, that does not mean that the individual making such comments has free rein to make them in any circumstance at all. The individual's freedom to express their views is limited to the extent provided for by art 10(2) and it will then be for the Court to assess whether any limitation imposed by the State is justified.'

51 [2000] HRLR 249 [20].

52 *Handyside v United Kingdom* (1979–80) 1 EHRR 737, [56]; *Sunday Times v United Kingdom (No 2)* (1992) 14 EHRR 123, [50].

53 *Re Christian Institute* [2007] NIQB 66.

54 Equality Act (Sexual Orientation) Regulations (Northern Ireland) 2006 (SI 2006/439).

> The belief in question is the orthodox Christian belief that the practice of homosexuality is sinful. The manifestation in question is by teaching, practice and observance to maintain the choice not to accept, endorse or encourage homosexuality. Whether the belief is to be accepted or rejected is not the issue. The belief is a long established part of the belief system of the world's major religions. *This is not a belief that is unworthy of recognition*. I am satisfied that Article 9 is engaged in the present case. The extent to which the manifestation of the belief may be limited is a different issue.[55]

More recently, in *Omooba v Michael Garret Associates*,[56] an actress had her offer of a role in a play where the character would show lesbianism in a sympathetic light withdrawn when her own negative views on homosexuality came to light. The Employment Tribunal dismissed her claim on finding that the offer was withdrawn not because of her views as such but because, when these views became known and she refused to (sufficiently) retract them publicly, she would not be credible in the role and the production itself was threatened. Nevertheless the Employment Tribunal, if with noticeable hesitancy, affirmed that a belief that no one is born gay (that God makes everyone non-gay and LGBT people choose their sexuality, contrary to God's word, and are sinful) was entitled to respect and protection under article 9:

> While wholly understanding why the statement of the claimant's beliefs was deeply offensive to people of same sex orientation, as well as to those of other orientations and none, we could not go so far as to say that merely stating the belief was not worthy of respect in a democratic society. It does not advocate harassment, although the belief may from time to time be expressed in ways that do amount to intimidation, nor that gays should not be employed, run businesses, be punished or shunned. She did not suggest conversion therapy, though that is underpinned by the belief that you are not born gay. A pluralist society must respect belief, however unacceptable to many people. There may be limits, for example to incitement to violent action, or setting restrictions on other people leading their lives, but as expressed by the claimant, she was not advocating any more than that other Christians must express their beliefs …
>
> After anxious and careful consideration we concluded that the Claimant's beliefs as manifested in the Facebook post, did scrape over the threshold for protection [under article 9], having regard to section 9(2).[57]

One of the clearest judicial affirmations of the right to hold and to express the view that marriage should not be extended to same-sex

55 *Re Christian Institute* (n 53 above) [50] (italics added).

56 *Omooba v Michael Garret Associates*, Employment Tribunal Case Number 2202946/19, 2602362/19, 17 February 2021.

57 Ibid [93]–[94] (the reference to 'section' 9(2) is a slip of the keyboard: throughout, the Tribunal were discussing art 9 of the ECHR).

couples is to be found in *Foster v Jessen*,[58] an action for defamation raised by Arlene Foster, previously First Minister of Northern Ireland, against a defendant who had accused her of being (*inter alia*) homophobic due to her opposition to same-sex marriage. McAlinden J said that Foster's 'traditional religious views which … are an important part of who she is, are views which she is entitled to hold and entitled to express … As Northern Ireland becomes a more secular society, there must be room or accommodation for an individual to hold such traditional religious views without being automatically classed as homophobic.'[59]

Now, the test for the legitimacy of limitations on opposition to same-sex marriage is not found in the parameters of the contested word 'homophobic' (which is not a legal term of art) but in a proportionality analysis that will assess the acceptability of any limitations to the rights in articles 9 and 10 in the light of the aims sought by these limitations. The European Court has more than once acknowledged that:

> the notion of 'respect' … is not clear cut, especially as far as positive obligations are concerned: having regard to the diversity of the practices followed and the situations obtaining in Contracting states, the notion's requirements will vary considerably from case to case and the margin of appreciation to be accorded to the authorities may be wider than that applied in other areas under the Convention.[60]

The appropriate level of respect to be shown, for reasons of democracy, to unpopular views may be gleaned from the way Lady Hale expressed herself in *R (Williamson) v Secretary of State for Education*. She said that 'a free and plural society must expect to tolerate all sorts of views which many, even most, find completely unacceptable'.[61] The language of toleration was again used by the Employment Appeal Tribunal in *Forstater v CGD Europe* when it said:

> the legal recognition of Civil Partnerships does not negate the right of a person to believe that marriage should only apply to heterosexual couples … [Such] beliefs may well be profoundly offensive and even distressing to many others, but they are beliefs that are and must be tolerated in a pluralist society.[62]

---

58 [2021] NIQB 56 [36].

59 This reflects the words of the US Supreme Court in *Obergefell v Hodges* 135 SCt 2584 (2015), 2602, [12]: 'Many who deem same-sex marriage to be wrong reach that conclusion based on decent and honorable religious or philosophical premises, and neither they nor their beliefs are disparaged here.'

60 *Goodwin v United Kingdom* (2002) 35 EHRR 18, [72]; *Hämäläinen v Finland* Appl 37359/09, 16 July 2014; *Fedotova v Russia* Appl No 40792/10, 13 July 2021, [45].

61 *Williamson* (n 34 above) [77].

62 *Forstater v CGD Europe* (n 50 above) [116].

To 'tolerate' unpalatable views means no more than that the law must permit them to be expressed without legal punishment and to require states to restrict their expression only in a manner that is proportionate to the aim of protecting the rights of others.[63]

BEYOND TOLERANCE?

So, democracy tolerates intolerance by allowing it to be manifested in words and deeds, but it does not encourage it, and the ECHR does not require states to offer intolerance any higher level of respect and protection than toleration. To put it another way, while states must allow the expression of controversial, upsetting and offensive views (within limits discussed above), there is no positive obligation on them to encourage or facilitate such expression, nor to confer some positive legal advantage from holding such views, nor to protect the believer from the social consequences of expressing their beliefs. This may be illustrated in the following cases.

In *Ladele v Islington Borough Council*,[64] a local authority marriage registrar considered same-sex relationships to be sinful in the eyes of her god: she held absolutely and irrevocably to the belief in heterosexual superiority. Her right to hold that view is absolute under article 9. Her right to express that view is protected by article 10, though that right may be restricted if the mode of expression incites violence or unrest or amounts to a hate crime. But Ms Ladele wanted more than simply the right to express her view peacefully: she wanted an employment benefit, in the form of being relieved from the obligation of registering civil partnerships for same-sex couples that would otherwise be part of her job. She was attempting to use her belief in the sinfulness of homosexuality (the superiority of heterosexuality) to give her an exemption from her employer's policy of providing the registration service they offered to the public in a non-discriminatory fashion. Both the English Court of Appeal and the European Court of Human Rights had little difficulty in rejecting her claim.[65] They would not allow her to use her article 9 right of freedom of thought, conscience and religion to gain the legal advantage of an exemption not open to those of her colleagues who did not share her belief in

63 The refusal to prosecute a homophobic joke made against a gay television personality was found by the European Court of Human Rights not to constitute a violation of either art 8 or art 14 because the balance between these articles and art 10 had been properly struck: *Sousa Goucha v Portugal* Appl 70434/12, 22 March 2016. But the refusal to prosecute calls for a gay couple to be killed did violate arts 8 and 14 in *Beizaras and Levickas v Lithuania* (2020) 71 EHRR 28.

64 [2009] EWCA Civ 1357.

65 Ibid; *Eweida v United Kingdom* (2013) 57 EHRR 8.

heterosexual superiority. The protection given to her by the ECHR was no more than the right to express her views: they were respected (protected) to that extent only.

In *R (Cornerstone (North East) Adoption and Fostering Services) v OFSTED*,[66] a Christian charity that offered adoption services failed in their challenge to a requirement by the regulatory body that they remove a condition that their prospective adopters 'refrain from homosexual behaviour'.[67] They were held entitled to hold, and indeed to expound, the view that children should always be brought up by opposite-sex couples, but they could not use that belief to give them an exemption from the non-discrimination provisions in the Equality Act 2010 beyond the religious exemptions that the Act explicitly contains.

A similar case is that of *R (Johns) v Derby City Council*[68] where a couple with strongly anti-LGBT (religiously inspired) views had their approval as short-term foster carers denied. The court held that the attitudes of potential foster carers to sexuality was a relevant consideration in the approval process. The couple, by seeking approval, were submitting themselves to the National Minimum Standards for Foster Carers, and they could not require the approving council to exempt them from the standard that required non-discrimination on the ground of sexual orientation just because of their beliefs.

And, in *Page v Lord Chancellor and Lord Chief Justice*,[69] a magistrate who had refused to sign adoption papers for a same-sex couple seeking (unopposed) to adopt a child, and confirmed to the press that his views remained unchanged, was held not to have been unlawfully victimised when he was dismissed from his positions. His right to express his understanding of Christian belief, that children should be brought up by a man and a woman, was protected by article 10; but he wanted more. He wanted to apply that belief to the legal questions that his position as a magistrate required him to answer. He was effectively asking for the right to operate the adoption process in a manner consistent with his own beliefs rather than consistent with the law of the land. No judge can be granted that request in a society founded on the rule of law.

These cases suggest that only a very modest level of respect needs to be shown to homophobic beliefs, but the respect inherent in tolerance

66 [2021] EWCA Civ 1390.

67 OFSTED's requirement that the agency remove the ban on LGBT people was accepted to be an infringement under art 9(1), but its aim of ensuring a diverse and inclusive body of prospective adopters was a legitimate aim and the requirement was a proportionate means of achieving that aim, which in any case reflected Parliament's 'specifically expressed will' in the Equality Act 2010.

68 [2011] EWHC 375 (Admin).

69 [2021] EWCA Civ 254; and related case *Page v NHS Trust Development Authority* [2021] EWCA Civ 255.

is not without value to those who hold and express these beliefs. In *R (Ngole) v University of Sheffield*[70] a student on a social work degree programme had posted on social media his belief in the sinfulness of homosexuality, as a consequence of which the university withdrew him from the programme. His application for judicial review of that decision was successful because the university had misapplied its own disciplinary proceedings by assuming that the student would act upon his beliefs and discriminate against LGBT clients. The student won his case because he was claiming no more than the right to speak, and the toleration to be shown to his speaking was the limit of the respect to be accorded to that right. The Court of Appeal made clear that had there been evidence that the student would manifest his beliefs through his social work practice the university would indeed have been entitled to withdraw him from the course. In other words, he would not be allowed to use his own religious beliefs to carve out an exemption from the general law requiring him not to discriminate in his provision of social work services.[71] The applicants in *Re Christian Institute's Application for Judicial Review*,[72] already discussed, were likewise successful because they too were not claiming any exemption from existing legal rules: they were simply exercising their right to argue that such exemptions should be written into the legal rules. Ms Ladele, Cornerstone (North East) Adoption and Fostering Services, Mr and Mrs Johns, and your man Page on the other hand, lost their claims because they were seeking exemptions from legal rules presently in force and applicable to everyone who did not subscribe to their belief in heterosexual superiority.[73]

---

70 [2019] EWCA Civ 1127. For comment and comparison with *Lee v Ashers Baking Company Ltd* see Steve Foster, 'Accommodating intolerant speech: religious free speech versus equality and diversity' (2019) European Human Rights Law Review 609. See also Foster, 'Accommodating intolerant speech: the decision in *Ngole v University of Sheffield*' (2020) 25 Coventry Law Journal 108.

71 A similar issue arose in Dundee Sheriff Court in *Keogh v University Court of Abertay University* 2023 GWD 2-20, 12 December 2022, where the university instituted a complaints procedure against a student who (other students felt) was expressing transphobic views in class, but dismissed the complaint. The student sued the university for subjecting her to the complaints procedure (which she accepted had been conducted properly) but failed on the ground that she had not been discriminated against: the complaints process had not been instituted because she had expressed views, but because other students complained about her doing so.

72 *Re Christian Institute* (n 53 above).

73 See also Stijn Smet who explains the decision in *Ladele* as the English court's resistance to allowing toleration to accommodate the 'expressive harm' that Ms Ladele's refusal to register civil partnerships necessarily caused, contrary to the 'expressive function' of the non-discrimination policy at issue: 'Conscientious objection to same-sex marriage: beyond the limits of toleration' (2016) 11 Religion and Human Rights 115.

Omooba v Michael Garret Associates,[74] the case of the actress whose offer of a role as a gay-friendly character in a play was withdrawn, illustrates the same point in a different way. Ms Omooba was not seeking an exemption from any rule of law or employment obligation, as Ms Ladele was, but she was asking the law to protect her from the economic consequences of expressing unpopular views, including from the unwillingness of others to maintain contractual arrangements with her. Her views were entitled to respect, to the minimal extent that the law had to tolerate her expressing them, but it was her own responsibility to balance her decision to express these views with the effect doing so would have on her future employment prospects. Effectively, she was asking for a higher level of protection: the right to express views and at the same time to have these views disregarded by others when they proved inconvenient to her. The Employment Tribunal held that she had only the right to express her views: it was up to her to take responsibility for the social and economic consequences of doing so.

GIVING ASHERS MORE

Applying these principles to the owners of the Ashers Baking Company Ltd, the law clearly and properly recognises that they have an absolute right to hold their belief in heterosexual superiority, and that they have a strongly protected right to manifest that belief by expressing opposition, in political debate, to any advancement of the legal rights of those they consider inferior to themselves, including marital rights for same-sex couples. But, like Ms Ladele and Mr Page, and unlike Mr Ngole, they demanded something more positive from the law than the right to believe themselves superior and indeed to say so. They demanded that the state give them benefit from that feeling of superiority in the form of an exemption from the non-discrimination rules contained in the Fair Employment and Treatment (Northern Ireland) Order 1998, which required everyone to provide the goods and services they offered to the public without discrimination on the ground of political belief. Like Ms Omooba they sought to be protected from the economic consequences of manifesting their beliefs – in her case losing a role in a play irrespective of her perceived suitability, in their case the obligation to pay damages for discrimination. The UK Supreme Court ought to have rejected their demand, just as the European Court rejected Ms Ladele's demand and Employment Tribunals rejected the demands of Ms Omooba and Mr Page. Its refusal to do so therefore offers higher protection than mere toleration of the defendants' belief in heterosexual superiority, a belief inconsistent with the values underpinning the ECHR. By allowing positive legal benefit

74 See *Omooba* (n 56 above).

to be gained from the defendants' belief in heterosexual superiority the Supreme Court gives it a legitimacy, an acceptability, a respectability, that runs perilously close to endorsing the belief itself.

CONCLUSION

There are two central flaws at the heart of the Supreme Court's decision in *Lee v Ashers Baking Company Ltd*. The first is that it affords the same level of respect to opposition to the opening of marriage to same-sex couples as it does to support for same-sex marriage, notwithstanding that opposition is inconsistent with, while support reflects, the values that underlie the ECHR. That led to the Supreme Court giving the belief in heterosexual superiority a level of protection greater than is justified by existing jurisprudence. The second central flaw is that this new level of respect is greater, almost certainly, than that which would be afforded to a belief in racial superiority, and the decision thereby creates a retrograde hierarchy of legitimacy where homophobia is less unacceptable (more acceptable) to the law than racism. This too is inconsistent with Strasbourg jurisprudence. While opposition to same-sex marriage is in itself neither a statement of hatred nor a call to violence, its common underpinning belief that heterosexuals are superior to homosexuals will, in Lord Dyson's words, 'encourage homophobic views'.[75] When the UK Supreme Court goes beyond toleration of such views, and affords them benefit, succour is given to those who do hate and who do wish to see that hatred manifested in violence and the violation of the rights of others. Until opposition to substantive LGBT equality is seen as exactly equivalent to opposition to racial equality, then the belief in heterosexual superiority will continue to be offered a greater level of respect and protection by the law than is consistent with the personal safety and emotional wellbeing of LGBT people. The decision serves to counter the message of moral equivalence between sexualities that hate crime prohibitions against homophobic speech and the legislation opening marriage to same-sex couples give[76] (and were designed to give) loud and clear. The losers, very directly, are LGBT people. Lord Dyson again: 'Homophobia places gays at risk.'[77]

---

75 *R (Core Issues Trust)* (n 38 above) [85].

76 See Kenneth Norrie, 'Now the dust has settled: the Marriage and Civil Partnership (Scotland) Act 2014' (2014) Juridical Review 135, 142–144.

77 *R (Core Issues Trust)* (n 38 above).

Northern Ireland
Legal Quarterly
Autumn Vol. 74.3 (2023) 439–471
Articles DOI: 10.53386/nilq.v74i3.1036

Rethinking the signature rule and the sufficiency of signatures as evidence of notice of terms in (business-to-business) commercial contracts

Moshood Abdussalam*

University of Otago, New Zealand
Correspondence email: moshood.abdussalam@otago.ac.nz

ABSTRACT

Using socio-economic analysis, this article critiques the oversimplification at the heart of the signature rule, which governs sufficiency in the notice of terms to a contractual counterparty concerning signed contracts in Commonwealth common law jurisdictions. As demonstrated in this article, two main factors account for the inadequacy of the signature rule as currently conceived. The first is the assumption that commercial entities are sophisticated. Second, in contested cases concerning notice, only manifest onerousness or unusualness of terms should warrant a heightened duty of notification on an offeror. This article argues that the signature rule lacks nuance and should be reformed to account for context-specificity. This is because: a) commercial sophistication is a matter of gradation; and b) terms need not be unusual or onerous to require heightened disclosure requirements – what matters for specific disclosure is the salience or peculiarity of a term.

Keywords: The rule in *L'Estrange v Graucob*/the signature rule; notice externalities; notice of contract terms; onerous terms.

INTRODUCTION: AN OVERVIEW OF THE ISSUES FOR DETERMINATION, METHOD AND THE THESIS PRESENTED

Consent is a fundamental factor whose absence generally deprives agreements of validity and enforceability. The rudimentary role of consent in the making of contracts tends to render new discussions about it to come across as hoary. While consent is not arcane, it can be complicated, given that consent is an elastic concept susceptible

* I would like to give thanks to Dr Jane Evans, Christopher Whitehead, Prof Charles Rickett, Prof Allan Beever and Prof Pey Woan for their feedback and helpful comments. I would also like to thank the anonymous reviewers for their constructive feedback.

to subtle manipulation.[1] The intricacies of modern commercial contracting demand a re-assessment of the rules and principles that govern consent, mainly as it concerns the role of signatures as evidence of notice of terms. This specific issue – the role of signatures as the manifestation of consent – is the focus of this article.

It is crucial to state five features of this article upfront. Firstly, this article pursues a prescriptive discussion rather than a descriptive discussion. Secondly, while this article draws on case law and legal academic discussions, it also relies on relevant basic socio-economic concepts. As the reader shall discover, the rationale for drawing on concepts of socioeconomics is that legal analysis alone does not adequately highlight the complexities of modern contracting. Thirdly, the article addresses only *business-to-business* (B2B) transactions; therefore, it excludes the treatment of *business-to-consumer* (B2C) transactions. The reason for this is that the interests of consumers in B2C transactions, as regards notice of terms, are well attended by different consumer protection regimes. For example, in the United Kingdom (UK), sections 62–69 of the Consumer Rights Act 2015 address the subject of notice or incorporation of terms. Fourthly, this article relates primarily to the incorporation of terms by signature and, as such, does not address the incorporation of terms by reference, nor does it address notice in cases of unsigned presentations of terms. It concerns those contracts whereby a signature is taken as a manifestation of consent and notice of terms. Fifthly, this article does not concern the regulation of (substantive) unfairness of contract terms, which is already dealt with by different statutes across various jurisdictions (eg the UK's Unfair Contract Terms Act 1977 and New Zealand's Fair Trade Act 1986). Instead, this article is concerned with the effective communication of contract terms in ways that do not create unfair surprises as regards the nature or scope of commercial risk or liability assumed by a counterparty.

The object and thesis pursued

This article critiques the oversimplification at the core of the signature rule, which governs notice or incorporation of terms in forming signed contracts in commercial settings. As this article demonstrates, the likelihood of weakened consent increases with the possibility of an offeror presenting important terms (ie those that ought to be salient terms) in unexceptional ways. The English High Court recently

1 Chunlin Leonhard, 'The unbearable lightness of consent in contract law' (2012) 63 Case Western Reserve Law Review 57; see also, Robin West, 'Authority, autonomy, and choice: the role of consent in the moral and political visions of Franz Kafka and Richard Posner' (1985) 99 Harvard Law Review 384.

expressed this concern in *Blu-Sky Solutions Ltd v Be Caring Ltd*.[2] In that case, HHJ Stephen Davies described the contested term(s) as 'concealed within detailed T&Cs, making it very hard to see the important from the unimportant'.[3] The concern is that offerees may be lured into contracts or ambushed by terms whose purpose they may not have understood to be salient, thereby causing them to discount their likely implications.

This article demonstrates two main flaws at the heart of the signature rule. The first is the specious assumption that business entities are (to be deemed) sophisticated. The essence of this assumption is that businesspersons have business knowledge and robust levels of preparedness for transactional risks (as compared to consumers). Therefore, by signing contracts, they agree to the terms contained therein. Contrary to this assumption, sophistication is often context-specific on a spectrum. The second flaw is the principle that only onerous or unusual terms must be brought to the attention of offerees for notice to be effective in contested cases regarding the sufficiency of notice. Doubts linger concerning such an exception on the reasoning that the quality of onerousness or unusualness of terms only applies to negate notice in cases of unsigned presentations of terms or consumer contracts.[4] Such a position is contestable, as several cases acknowledge that the onerousness of terms may negate notice.[5] In any case, this limited exception to the signature rule is problematic, as this article demonstrates. Of course, one may point to case law decisions whose *ratios* correspond with the two considerations this article advances. But, then, a legion of cases continues to strengthen the unvarnished predominance of the signature rule, therefore calling into question the reliability of those other cases that otherwise correspond to the above-stated considerations. On that account, this article seeks to provide theoretical reinforcement for advancing these considerations in judicial analysis.

This article deploys socio-economic analysis to critique the predominance of the signature rule, mainly from the standpoint of contracting processes. Such a process-based approach is not novel in contract law scholarship. It has been exponentially used to analyse

2 [2021] EWHC 2619.

3 Ibid para 112.

4 See David Foxton, 'The boilerplate and bespoke: should differences in the quality of consent influence the construction and application of commercial contracts?' in Charles Mitchell and Stephen Watterson (eds), *The World of Maritime and Commercial Law: Essays in Honour of Francis Rose* (Hart 2020) 259, 261–265.

5 See, cases in footnotes 12 and 13 below.

or deconstruct judicial interpretation of contracts[6] and rationales that inform conceptions of contract terms, and so on.[7] But the novelty of its deployment, as used in this article, is deciphering how contracting patterns or processes may impact the effectiveness of communicating contract terms to counterparties. Central to its novelty is the postulation that judges be guided by two parameters that serve as essential touchstones for effective communication through a textual medium when ascertaining whether notice is sufficient. These are the factors of *relational proximity* and *informational complexity*. With these parameters, we can objectively determine what terms parties consider salient, distinguished from those that qualify as non-salient. Also, with these parameters, we can discern what terms parties to a contract ought to have reasonably expected or, at least, be on guard about when forming contracts. Building on these parameters, the main submissions of this article are as follows:

a) the signature rule should generally apply to negotiated and industry-standard terms; and
b) the signature rule should not generally apply to unilaterally dictated terms.

Structure of the article

Section two of this article highlights the formalist foundations of the signature rule. It also emphasises the importance of notice as an element of consent, particularly that (insufficient) notice can have severe private and social implications. Section three addresses the complexities of contractual consent in the contemporary marketplace. Here, the difficulties of applying the signature rule in the contemporary marketplace to contracts that are not outcomes of negotiations are highlighted. These intricacies primarily arise from informational complexity and tenuous relational proximity. With these intricacies come the possibilities of manipulating an offeree‘s consent to express terms. The fourth section addresses the assumption that business entities are sophisticated and should therefore be assumed to have notice of all express terms contained in contract documents. Section

6 John F Coyle and W Mark C Weidemaier, ‘Interpreting contracts without context’ (2018) 67 American University Law Review 1673, 1677–1678; see also Stephen J Choi, Mitu Gulati and Robert E Scott, ‘The black hole problem in commercial boilerplate’ (2017) 67 Duke Law Journal 1; see also, Ronald J Gilson, Charles F Sabel and Robert E Scott, ‘Text and context: contract interpretation as contract design’ (2014) 100 Cornell Law Review 23–97.

7 Ronald J Gilson, Charles F Sabel and Robert E Scott, ‘Contracting for innovation: vertical disintegration and interfirm collaboration’ (2009) 109 Columbia Law Review 431; Matthew Jennejohn, ‘Do networks govern contracts’ (2022) 47 Journal of Corporation Law 333–386.

five critiques the requirement that only onerous and unusual terms should be specifically brought to the notice of offerees. The final section provides recommendations for improving the rules on notice and concludes the article.

THE IMPERATIVE TO REVISIT NOTICE AS AN ELEMENT IN THE CONSENT MATRIX

The signature rule and its formalist foundations

In common law jurisdictions, the signature rule is the starting point for determining consent to signed express contractual terms in commercial settings.[8] A signature to express terms is taken as bearing valuable evidentiary weight against a party who signed a contract. As Lord Denning described it in *Curtis v Chemical Cleaning & Dyeing Co*, 'If the party affected signs a written document, knowing it to be a contract which governs the relations between them, his signature is irrefragable evidence of his assent to the whole contract.' Similarly, in *Harris v Great Western Railway Co*,[9] Blackburn J (as he was then) described the force of signatures, saying:

> I apprehend … that, by assenting to the contract thus reduced to writing, he represents to the other side that he has made himself acquainted with the contents of that writing and assents to them, and so induces the other side to act upon that representation by entering into the contract with him.[10]

According to this rule that originates from the decision in *L'Estrange v Graucob*,[11] once contractual counterparties sign commercial documents, it is taken that they have consented to the terms contained therein. This is so even if specific terms are contained in the document whose presence they have only discovered later.[12] They probably would not have entered the contract or sought to negotiate around them if they had known such terms beforehand. However, the rule does not apply where contract documents are unsigned. As this article focuses solely on signed contracts, the rules and principles that apply to unsigned contracts shall not be addressed.

Apart from agreements tainted with vitiating factors (eg fraud, misrepresentation, undue influence, or mistake), there is a heavily

8 Bruce Clarke and Stephen Kapnoullas, 'When is a signed document contractual – taking the fun out of the funfair' (2001) 1 Queensland University of Technology Law and Justice Journal 39; Matthew Chapman, 'Common law contract and consent: signature and objectivity' (1998) 49 Northern Ireland Legal Quarterly 363.

9 (1876) 1 QBD 515.

10 Ibid 530.

11 [1934] 2 KB 394.

12 *Bedford Investments Ltd v Sellman* [2021] EWHC 799 (Comm), para 65.

contested exception to the signature rule. That exception supposedly arises where the contract contains onerous or unusual terms.[13] But the viability of that exception is doubtful in English law, as its soundness has been questioned by modern cases that have ruled that mere onerousness is insufficient; rather, the import of the terms must be (evidently) extortionate.[14] McMeel labelled the English judicial philosophy of enforcing contractual agreements at face value as informed by 'documentary fundamentalism'.[15] A similar approach prevails in Australia, where governing legal authority considers a signature evidential of consent, except if the signatory was induced into the contract (or into signing) by improper or inequitable means.[16] The same judicial philosophy dominates in New Zealand,[17] Northern Ireland[18] and the Republic of Ireland,[19] to mention a few common law jurisdictions. But a different approach appears to prevail in Scotland,[20] where insufficient disclosure of the onerousness of terms may render a counterparty's signature an inadequate representation of their consent.

As Miller observes, the signature rule is a species of legal formalism that courts raise as justification for non-interference in the regulation of term incorporation because signatures represent autonomy and informed consent.[21] The attribution of autonomy and informed consent to businesspersons is (subtly) rooted in the assumption that businesspersons are sophisticated. Miller aptly describes this

13 *One World Ltd v Elite Mobile Ltd* [2012] EWHC 3706 (QB), per HHJ Behrens: 'I am content to assume (without deciding) that there is a possible exception to the rule in L'Estrange v F Graucob Ltd in relation to provisions that are onerous or unusual.' See also, Elisabeth Peden and J W Carter, 'Incorporation of terms by signature: L'Estrange Rules!' (2005) 21 Journal of Contract Law 96.

14 *DO-BUY 925 Ltd v National Westminster Bank plc* [2010] EWHC 2862 (QB); *Cargill International Trading v Uttam Galva* [2019] EWHC 476 (Comm), paras 79–94; *Woodeson & Another v Credit Suisse (UK) Ltd* [2018] EWCA Civ 1103, para 42; *Higgins & Co Lawyers Ltd v Evans* [2019] EWHC 2809, para 73.

15 Gerard McMeel, 'Documentary fundamentalism in the senior courts: the myth of contractual estoppel' (2011) Lloyd's Maritime and Commercial Law Quarterly 185.

16 *Fox Tucker Pty Ltd v Morgan* [2023] SASCA 11, para 57; *Toll (FGCT) Pty Ltd v Alphapharm Pty Ltd* [2004] HCA 52; *Ozmen Entertainment Pty Ltd v Neptune Hospitality Pty Ltd* [2019] FCA 721.

17 *Whitley v Ribble Ltd* [2017] NZHC 1884; *Cygnet Farms Ltd v ANZ Bank New Zealand Ltd (No 2)* [2017] NZCCLR 4.

18 *Lambe v AIB Group (UK) plc* [2020] NIJB 497; *Ulster Bank Ltd v Taggart* [2011] NIMaster 1.

19 *James Elliott Construction Ltd v Irish Asphalt Ltd* [2014] IESC 74.

20 See *Montgomery Litho Ltd v Maxwell* 2000 SC 56; *Difference Corporation Ltd v Unitel Direct Ltd* [2019] SC EDIN 56; *Brandon Hire plc v Steven Russell* [2010] CSIH 76.

21 Meredith Miller, 'Contract law, party sophistication, and the new formalism' (2010) Missouri Law Review 494.

assumption as based on the view that businesspersons 'have access to information, resources to allocate risk and experience or predisposition to counteract cognitive bias'.[22] In other words, sophistication connotes relevant transactional expertise or acquaintance. This reasonably explains the law's (generally) non-interventionist approach in B2B transactions compared to B2C transactions.[23] In transactions involving consumers, there is a foundational view that consumers tend to lack sophistication based on the assumption that they are unlikely to possess an awareness of the complexities and profundity of details that shape relevant commercial bargains, contexts, or environments.[24]

Two main rationales inform the signature rule. These are rationales often advanced by proponents of contractual formalism.[25] The first is the objective assessment of contractual consent, which states that to determine consent we must look at the outward manifestations of parties to a contract, as we cannot judge their subjective intentions.[26] The second is efficiency and certainty in commercial arrangements.[27] That the rule allows immediate parties to agreements and connected third parties to rely on contractual documents, taking consent as a settled matter. McLean justifies the signature rule as ensuring that businesspeople act with more prudence and caution when entering contracts.[28] It ensures that they take time to peruse documents before signing them. This rationale pivots on the assumption that businesspersons are sophisticated.[29] While the need for certainty in commercial life is understandable, the signature rule appears too formulaic and rigid to guide the ascertainment of notice of express terms. Sadly, courts applying the signature rule ignore the reality that

22 Ibid 495.

23 See *Bankway Properties Ltd v Penfold-Dunsford* [2001] 2 EGLR 36, para 40; see also *Carnival plc v Karpik (The Ruby Princess)* [2022] FCAFC 149, at paras 193, 206(d).

24 See Larry T Garvin, 'Small business and the false dichotomies of contract law' (2005) 40 Wake Forest Law Review 295.

25 See Jonathan Morgan, *Contract Law Minimalism: A Formalist Restatement of Commercial Contract Law* (Cambridge University Press 2013) 223–224; see also Alan Schwartz and Robert Scott, 'Contract interpretation redux' (2010) 119 Yale Law Journal 926.

26 Matthew Chapman, 'Common law contract and consent: signature and objectivity' (1998) 49 Northern Ireland Legal Quarterly 363; see also R A Samek, 'The objective theory of contract and the rule in L'Estrange v Graucob' (1974) 52 Canadian Bar Review 351.

27 See, *Codelfa Construction Pty Ltd v State Rail Authority of New South Wales* [1982] HCA 24; see also *Platform Funding Ltd v Bank of Scotland plc* [2009] 1 QB 426.

28 Hazel McLean, 'Incorporation of onerous or unusual terms' (1988) Cambridge Law Journal 172, 174;

29 See, Schwartz and Scott (n 25 above).

an inquiry as to whether a party had notice of terms, even in cases of signed documents, requires a context-specific assessment.

Understandably, heightened disclosure or notice requirements may be useless.[30] What is the point in requiring an offeror to make adequate disclosure of a term that the offeree ought reasonably to have expected as customary in a type of contract? Of what need is heightened disclosure of terms when the cost of doing so may overshadow the joint value of the transaction to the counterparties? In other words, what is the point of disclosure when it possibly exceeds the value of a given transaction? We can adapt the example supplied by Fairfield to illustrate these questions.[31]

Fairfield suggests we assume a high-volume but low-value business operated by an offeror, such as a coffee shop. Suppose each cup of coffee sold generates, on average, $1 of joint gain to the counterparties. If the law required the shop operator to specifically bring it to the notice of each patron that the shop operator shall not be liable for burns suffered by patrons while consuming hot coffee, such law would be of no good to both the patrons and the operator. It would delay transactions and increase the operator's operating cost in ways that may not be well compensated by the competitive price charged for each cup of coffee. It also adds no value to the interest of patrons as it assails them with information that they should know or are most likely to be uninterested in knowing.[32] However, disclosure is necessary for high-value transactions, which B2B bargains commonly are, where surprise can be costly to the offeree.[33] Fairfield says 'preventing contractual surprise may be worth the cost' concerning such cases.[34] Thus, we now discuss the need for heightened notice requirements in commercial settings.

The cost of notice failures

Lack of notice in commercial transactions is of particular interest because of the propensity to create notice externalities, which may impose a cost on offerees and society in aggregate. As reasoned by Menell and Meurer, notice failures resulting from the ineffective communication of an interest owned by a party to other persons likely to encounter the interest may impose both private and social costs.[35]

30 See Carl Schneider and Omri Ben-Shahar, *More Than You Wanted to Know: The Failure of Mandated Disclosure* (Princeton University Press 2016).

31 Joshua Fairfield, 'The cost of consent: optimal standardization in the law of contract' (2009) Emory Law Journal 1402, 1408.

32 Ibid 1423–1426.

33 Ibid.

34 Ibid 1425.

35 Peter Menell and Michael Meurer, 'Notice failure and notice externalities' (2013) 5 Journal of Legal Analysis 1.

Such notice failures are likely to secure opportunistic gains for a party at the expense of other persons who were not effectively aware of the said interest beforehand, and persons connected to those other persons. We can refer to the former cost species as micro-level/interpersonal effects and the latter as macro-level/social effects.

Micro-level or interpersonal effects may arise where an offeror can impose vital terms in ways that bypass the awareness of an offeree. This is because there would be a negation of genuine consent as the offeree is not allowed a fair opportunity to fully consider his assumption of contractual relations with the offeror. As Murray aptly describes it, just as, 'government must correctly design and communicate its actions so as to offer the benefit or impose the tax that it intends to bestow upon its audience of citizens',[36] in the same vein, 'private actors must correctly design and communicate their actions to accurately offer the intended inducement or impose the intended price or rent'.[37] In essence, when people are effectively or sufficiently aware of what they are getting into, we can fairly say that they have exercised their choice regarding their private assessment of costs and benefits. Or, at least, they have been given a fair opportunity to assess their options rationally. Where notice of vital contractual terms is ineffective, offerees would fall into misperception problems, creating room for transactions that allow distributive disparities between an offeror and offeree.[38] Thus, the former will likely secure gains at the latter's cost.

For example, in *Blu-Sky Solutions Ltd v Be Caring Ltd*, the judge reasoned that the actual loss likely to be suffered by the offeror from the offeree's breach was less than 13 per cent of the sum of £180,000 that the offeror sought to claim based on the remedial clause inserted in the contract.[39] Such unfair distribution of gains accruable to the offeror enables undue economic rent.[40] It must also be noted that undue economic rent resulting from notice failure is not limited to price effects. Undue rent may take non-price forms such as the forbearance, liabilities, or risks that an entity transfers to other counterparties. That is, the burden transferred to the offeree exceeds the value supplied by the offeror towards the bargain between them. As shall become evident, this article is mainly concerned with effectively communicating non-price terms.

36 Michael D Murray, 'The great recession and the rhetorical canons of law and economics' (2012) 58 Loyola Law Review 615, 643.

37 Ibid.

38 Oren Bar-Gill, 'Algorithmic price discrimination: when demand is a function of both preferences and (mis)perceptions' (2019) 86 University of Chicago Law Review 217.

39 *Blu-Sky Solutions* (n 2 above) para 107.

40 Mariana Mazzucato, *The Value of Everything: Making and Taking the Global Economy* 1st edn (Public Affairs 2018) ch 7.

We now shift to the macro-level effects of notice failures. Whittaker laments the common law's limited advertence to social interests when regulating contractual relations.[41] Most pertinently, he re-echoes the fears of the UK Competition Authority that the transfer of excessive or unexpected burdens by large retailers to their suppliers risks the propensity for a diminution in the incentives of those suppliers to invest in 'new capacity, products and production processes'.[42] The actual sufferers of such outcomes will ultimately be consumers who bear the brunt of higher living costs. This fear has already manifested itself in unfavourable welfare outcomes in society. Studies show a squeeze in living standards, bleak social mobility prospects, and high corporate and household indebtedness levels, among other dismal conditions.[43] Much of these outcomes are attributable to widening economic inequality gaps perpetuated by market concentration and power gaps. The higher the possibility of a firm possessing market powers to transfer unexpected economic burdens and costs to several counterparties, the higher the likelihood we would experience adverse social effects.[44] The concern becomes starker when offerors with market powers can subtly transfer burdens to counterparties. The consequence of this is the heightening of the cost structure of businesses across the economy.[45] Firms that have burdens transferred to them would have to shift the incidence of those costs to their customers, and their customers, to other persons down the supply chain.

THE NUANCES OF CONTRACTUAL CONSENT IN THE CONTEMPORARY MARKETPLACE

This section lays the foundation for the arguments that commercial sophistication is a matter of gradation and that terms need not be onerous or unusual to require heightened notice requirements. Pursuing this objective, this section highlights the nuances and peculiarities of different models or patterns of contracting that reflect

41 Simon Whittaker, 'Unfair terms in commercial contracts and the two laws of competition: French law and English law contrasted' (2019) 39 Oxford Journal of Legal Studies 404.

42 Ibid 428.

43 Brian Nolan, Matteo Richiardi and Luis Valenzuela, 'The drivers of income inequality in rich countries' (2019) 33 Journal of Economic Surveys 1285; see also, Atif Mian, Ludwig Straub and Amir Sufi, 'Indebted demand' (2021) Quarterly Journal of Economics 2243-2307.

44 See Kathleen Engel, 'Do cities have standing? Redressing the externalities of predatory lending' (2006) 38 Connecticut Law Review 355.

45 See Mariana Mazzucato, Josh Ryan-Collins and Giorgos Gouzoulis, 'Theorising and mapping modern economic rents' UCL Institute for Innovation and Public Purpose, Working Paper 2020.

the difficulties of a general or literal application of the signature rule. A two-step process shall be followed to establish the need for heightened disclosure in commercial settings. The first step shows that contracting is mainly conducted through terms that are not negotiated. The second is to demonstrate that, as contracts are increasingly outcomes of terms that are not negotiated, there would unavoidably be significant communication gaps between counterparties. The purpose of this exercise is to advance the case of this article for tailoring notice requirements to contextual needs.

Peculiarities of contracting processes in modern commercial settings

In theory, contracts between business entities are outcomes of negotiations.[46] But, as reality has it, commercial contracts between business entities usually result from three main contracting methods: *negotiations*, *industry standards* and *unilateral dictation of terms*. On the one hand, these three methods of contractual formation may be informed by conditions of competition in a market or a relative balance in bargaining powers between parties and, on the other, by factors of transactional convenience or transaction costs. Parties negotiate contracts where they have counterbalanced bargaining powers or because of the relative competitiveness of the market. Such situations prevent one party from dictating terms to the other without the other countering the terms proposed or shifting to alternative offerors in the market. The possibility of parties to freely negotiate terms based on relatively balanced bargaining positions is unusual in the modern marketplace. This is because negotiated contracts are mostly the reserve of parties with relatively equal bargaining powers.

An excellent example of this is merger and acquisition agreements. Most transactions in the modern economy are not negotiated; they are products of industry standards and unilaterally dictated terms. Confusingly, both contracting patterns are referred to as standard-form contracts. On account of the pervasiveness of these two patterns, Robertson suggests that 'contract scholarship must therefore take the standard form, rather than the negotiated transaction, as its central focus'.[47]

Before explaining how these three contracting methods impact effectiveness in the communication of terms, terminological clarification on the meaning of 'standard form/terms' (also called boilerplate terms) is essential. The terminology is commonly (but inaptly) used to

46 Peter Benson, *Justice in Transactions: A Theory of Contract Law* (Harvard University Press 2019) 217.

47 Andrew Robertson, 'The limits of voluntariness in contract' (2005) 29 Melbourne University Law Review 179.

connote industry standards and unilaterally dictated terms. A classic example of such usage is in *Macaulay v Schroeder Music Publishing Co Ltd*,[48] where Lord Diplock asserted that 'standard form of contracts are of two kinds'. The first are 'those which set out the terms upon which mercantile transactions of common occurrence are to be carried out'. He cited as examples 'bills of lading, charterparties, policies of insurance, contracts of sale in the commodity markets'. He described this first kind as involving terms that 'have been settled over the years by negotiation by representatives of the commercial interests involved and have been widely adopted because experience has shown that they facilitate the conduct of trade'.[49] He rightly describes the second kind of standard form as the result of 'the concentration of particular kinds of business in relatively few hands'.[50] They involve the use of terms dictated by one party to other persons (ie the offerees), as the terms 'have not been the subject of negotiation between the parties to it, or approved by any organisation representing the interests of the weaker party'.[51]

We find both judges[52] and reputable scholars using the term 'standard form' without clear distinction between both contracting methods.[53] It is submitted that both methods of contracting should be distinguished.[54] The first may be known as industry-standard terms, while the second may be regarded as unilaterally dictated terms.[55]

Industry standard terms

When parties enter contracts following industry standards, they agree to opt into a regime of terms in a manner analogous to a *plug-and-play* system. Generally, the use of industry standards is informed by transaction cost considerations and predictability in meaning and

48 [1974] 1 WLR 1308.

49 Ibid

50 Ibid

51 Ibid

52 See, for example, the UK Supreme Court decision in *Triple Point Technology, Inc v PTT Public Company Ltd* [2021] UKSC 29; see also the Canadian Supreme Court decisions in *Ledcor Construction Ltd v Northbridge Indemnity Insurance Company* 2016 SCC 37 and *Uber Technologies Inc v Heller* 2020 SCC 16.

53 See, for example, Margaret Jane Radin, *Boilerplate: The Fine Print, Vanishing Rights, and the Rule of Law* (Princeton University Press 2014); Gregory Klass, 'Boilerplate and party intent' (2019) 82 Law and Contemporary Problems 105; Douglas G Baird, 'The boilerplate puzzle' (2006) 104(5) Michigan Law Review 933.

54 See, for example, Robert Merkin and Jenny Steele, *Insurance and the Law of Obligations* (Oxford University Press 2013) 47; see also, Aaron Taylor, 'What is a standard form?' (2017) 33 Professional Negligence 268.

55 Richard A Posner and Lucian A Bebchuk, 'One-sided contracts in competitive consumer markets' (2006) 104 Michigan Law Review 827.

commercial expectations. Participants within close-knit commercial communities or specialised aspects of commerce often adopt such terms. The governance mechanisms of relevant trading platforms, societies, associations, or networks standardise the terms.[56] Such terms are often prepared and systematised in a 'hub-and-spoke' fashion. Various examples of such contracting regimes abound in commercial life.

A good example is the International Swaps and Derivatives Association master agreement (ISDA) regime which standardises contract terms for transactions relating to over-the-counter derivatives in financial markets.[57] Another is the New York Produce Exchange Form (NYPE), the most used standard terms for time charterparties in the shipping industry.[58] Further examples can be found in various Technology Standards Licensing Agreements.[59] When using this route, dickering over the generality of the contractual terms is eliminated. Transacting parties may only negotiate on limited and salient terms peculiar to their particular or contextual needs, such as price, duration and quantity.[60] In effect, such a contracting route enables business entities to have predictability and certainty in their expectations, especially when the interpretation of major terms has been subject to judicial determination.[61]

Given the apparent dynamics of this contracting route, one may view that parties adopting such contract terms are often well matched in terms of resources and operate under relatively competitive conditions, as no one may unilaterally dictate terms to counterparties. But this is not necessarily so. Some business entities in an industry where such standard terms are used may have weaker bargaining

---

56 See Ross Cranston, *Making Commercial Law through Practice 1830–1970* (Cambridge University Press 2021) 5–12.

57 Joanne Braithwaite, *The Financial Courts: Adjudicating Disputes in Derivatives Markets* (Cambridge University Press 2021) 31–49; see also, Sean Flanagan, 'The rise of a trade association: group interactions within the International Swaps and Derivatives Association' (2001) 6 Harvard Negotiation Law Review 21.

58 Johanna Hjalmarsson, 'Trip charterparties and their binary endgames' (2018) Lloyd's Maritime and Commercial Law Quarterly 376, 379; see also, Baris Soyer and Theodora Nikaki, 'Enhancing standardisation and legal certainty through standard charterparty contracts: the NYPE 2015 experience' in Barış Soyer and Andrew Tettenborn (eds), *Charterparties: Law, Practice and Emerging Legal Issues* (Routledge 2017) 67–89.

59 See Igor Nikolic, *Licensing Standard Essential Patents: FRAND and the Internet of Things* (Hart 2021); see also, Gregory Sidak, 'The FRAND contract' (2018) 3 Criterion Journal on Innovation 1.

60 *Homburg Houtimport BV v Agrosin Private Ltd (The Starsin)* [2003] 2 WLR 711, 777.

61 *Beaufort Developments (NI) Ltd v Gilbert-Ash NI Ltd* [1999] 1 AC 266, 274.

powers than others (possibly a few) who are better resourced.[62] An excellent example of such a situation is in the Australian construction industry. Contracting in that industry is likened to a pyramid structure whereby head contractors, usually few, can use industry standards to lock sub-contractors into unfavourable terms.[63] But the unfairness (or otherwise) of terms, *per se*, is not this article's focus. Since the focus is on notice, our concern is the use of 'transactional expertise' as a valuable indicator of notice. It is fair to say that persons constantly engaged in a particular industry can be regarded as experts, and such notice of standard terms in their industry can be attributed to such persons. Whether they are big or small players, robustly resourced or meagrely, this is so.

Unilaterally dictated terms

We come to the deployment of unilateral dictation of terms in commercial transactions. By unilateral terms, we mean contracts whose terms are dictated by one party to counterparties on take-it-or-leave-it conditions. Terms may be offered on take-it-or-leave-it conditions, yet offerees or counterparties may engage the offeror to revise the terms individually. Such situations involve converting what would otherwise have been unilateral dictations into negotiated contracts, as the offeree would have to make counteroffers, some forbearance, or provide a reason for the offeror to revise the terms. Unilateral terms are usually outcomes of in-house conception and drafting.

In some cases, they may be products of adaptations from, or direct duplications of, relevant industry standards or other pre-existing unilateral terms used by other commercial entities.[64] Most contracts are conducted based on unilaterally dictated terms, whether in matters of transportation, commercial lease, equipment purchase or leasing, insurance, hospitality, digital and telecommunications services, or outsourcing.[65] In financial markets, we find examples of this pattern

62 Vicki Waye and Jeremy Coggins, 'Squeezing out the market for lemons: the case for extending unfair contract terms regulation in the commercial context' (2020) 36 Journal of Contract Law 230.

63 Parliament of Australia, Senate Economic References Committee, *Insolvency in the Australian Construction Industry* (Commonwealth of Australia, Canberra 2015) 12–14.

64 W Mark C Weidemaier, 'Disputing boilerplate' (2009) 82 Temple Law Review 1, 2.

65 See, generally, Jeffrey W Stempel, 'The insurance policy as thing' (2009) 44 Tort Trial and Insurance Practice Law Journal 813; Lisa Bernstein, 'Beyond relational contracts: social capital and network governance in procurement contracts' (2016) 7 Journal of Legal Studies 561; Omri Ben-Shahar and James White, 'Boilerplate and economic power in auto manufacturing contracts' (2006) 104 Michigan Law Review 953.

used in arrangements concerning corporate and sovereign bonds, syndicated loans, and collateralised debt obligations.[66]

Unilateral dictation of terms in commercial settings is often enabled by market structure effects and competitive conditions in a relevant market or transactional convenience and costs issues. Where offerees with relatively balanced or even stronger bargaining positions to offerors accept unilateral terms without negotiating them, then it is fair to assume that they did so because of transactional convenience or because they consider the terms tolerable from among other alternative market offerings. In most cases, unilateral terms are usually accepted without negotiation because of the term-giver's (or offeror's) relatively superior market or bargaining position over the term-takers (or offerees). Also, it is not uncommon for dictated terms to be imposed on another firm of (relatively) equal bargaining powers following a battle of forms between the firms.[67]

Under imperfectly competitive market conditions, we can expect that, although markets would not be perfectly competitive, business entities that secure market lead would leave significant space for other competitors to exist and operate. Unfortunately, in modern times, we are confronted with conditions of market concentration or (almost) winners-take-all states occasioned by oligopolistic and oligopsonistic conditions.[68] Studies confirm that a blend of factors has endowed certain firms with degrees of market power (usually of oligopolistic or oligopsonistic quality) that other firms cannot easily measure up to or approximate. Such factors include mergers and acquisitions; market connections; economies of scale; strong brand loyalty; cost-cutting advantages; technological superiority; a web of intellectual property rights and other intangible assets; and superb managerial capabilities.[69]

---

66 See, generally, Anna Gelpern, Mitu Gulati and Jeromin Zettelmeyer, 'If boilerplate could talk: the work of standard terms in sovereign bond contracts' (2019) 44 Law and Social Inquiry 617; see also Şenay Agca and Saiyid Islam, 'Securitised debt markets' in H Kent Baker, Greg Filbeck and Andrew Spieler (eds), *Debt Markets and Investments* (Oxford University Press 2019) 131–148; see also Marcel Kahan and Mitu Gulati, 'Contracts of inattention' (2021) 46 Law and Social Inquiry 1115.

67 *Transformers & Rectifiers Ltd v Needs Ltd* [2015] BLR 336.

68 See Chris Carr, *Global Oligopoly: A Key Idea for Business and Society* (Taylor & Francis 2020); see also Luis Suarez-Villa, *Technology and Oligopoly Capitalism* (Routledge 2023).

69 Jan Eeckhout, *The Profit Paradox: How Thriving Firms Threaten the Future of Work* (Princeton University Press 2021) 52.

Firms endowed with these factors witness massive productivity levels relative to their competitors.[70] The upshot is that these firms secure a significant share of economic output, commercial transactions and profitability. For these reasons, such firms can expend more on overheads than other firms in the market. Still, they can also charge a higher markup in excess of their marginal cost of production (ie they can charge prices that compensate them for more than their cost of producing additional units of goods or services).[71] In comparison, their competitors would only struggle to maintain favourable levels of profitability and compete with these other 'superstar' firms in terms of price, costs, or quality.

The deployment of unilateral terms is not exclusive to oligopolistic/oligopsonistic firms. And as such, it is not entirely a market structure concern. The use of unilateral terms is pervasive in the economy as it is deployed by almost all small and medium-sized firms, imitating the practices of more prominent firms. Like prominent firms, these smaller firms tailor their terms towards improving their competitive positions. By presenting salient (non-price) terms as unimportant, they tend to lull offerees into discounting or ignoring the actual private cost of such terms, thereby increasing the market traction of their offers. Therefore, the imperative for heightened notice requirements in cases of term dictations is not necessitated only by market structure issues but also its pervasive use by commercial entities across the economy.

The intricacies of effectively communicating contract terms

Having identified that much of contracting in modern times is an outcome of both industrial standards and unilateral dictations, we now discuss how these contracting methods can impact the effective communication of terms. In determining the effectiveness of textual communication, two vital parameters to bear in mind are *information complexity* and *relational proximity*. We shall start by describing each of these parameters before discussing their implications on the effective communication of terms in situations of industry standards and unilateral terms.

Informational complexity relates to the degree of difficulty (or otherwise) with which the audience of a text may understand its contents. This parameter is primarily multidimensional and has

70 Sharat Ganapati, 'Growing oligopolies, prices, output, and productivity' (2021) 13 American Economic Journal: Microeconomics 309.

71 Jan De Loecker, Jan Eeckhout and Gabriel Unger, 'The rise of market power and the macroeconomic implications' (2020) 135 Quarterly Journal of Economics 561.

objective and subjective dimensions.[72] An objective indicator of information complexity is relatively measurable. Examples include the length of contract documents;[73] readability (eg whether written in fine print); and structural organisation (eg whether the document is divided into themes or headings).[74] However, subjective indicators relate to the degree of mental effort addressees require to decipher the meaning of texts. One good example is cognitive overload, which relates to an addressee's quantum of mental efforts to process and understand a body of text.[75] Another is bounded rationality, which involves addressees acting non-rationally by relying on biases and heuristics to discern the meaning of texts.[76] As scholars observe, informational complexity can pose significant challenges to addressees in knowing or understanding the content of a contract document.

Relational proximity, as a parameter, relates to the degree of connection and familiarity that a set of communication audiences share concerning each other's expectations.[77] Examples of parties sharing relational proximity include those with a history of dealings through past transactions. Such parties are commonly engaged in a particular industry and conversant with the customs and risk dynamics of the relevant industry. Persons sharing relational proximity include those who have met to negotiate the terms they intend to transact. Primary to relational proximity is the context against which we may understand the expectations objectively imputed to transacting parties when forming the contract. It becomes clear that, as relational proximity between parties becomes tenuous, the possibility of finding common expectations becomes more difficult. Tenuous relational proximity may arise from numerous conditions, such as the parties not having a history of previous dealings or that the parties barely or never met to discuss the terms of their contract.

---

72 John Hagedoorn and Geerte Hesen, 'Contractual complexity and the cognitive load of R&D alliance contracts' (2009) 6 Journal of Empirical Legal Studies 818, 821–824.

73 See David T Robinson and Toby E Stuart, 'Financial contracting in biotech strategic alliances' (2007) 50 Journal of Law and Economics 559.

74 Tal Kastner, 'Systemic risk of contract' (2022) 47 Brigham Young University Law Review.

75 Hagedoorn and Hesen (n 72 above) 825: 'Cognitive load also represents something more, for example, "a detailed schedule of payment amounts … will be more difficult to understand than a simple payment formula (for example, a 25% commission). And a payment of $X per widget will impose less cognitive load than an otherwise identical contract that bases payment on a fraction of profits which may be difficult to calculate".'

76 Melvin Eisenberg, 'The limits of cognition and the limits of contract' (1995) 47 Stanford Law Journal 211, 214.

77 Mark D Janis and Timothy R Holbrook, 'Patent law's audience' (2012) 97 Minnesota Law Review 72, 80–82.

From the description of both parameters, one can glean that they are connected. Relational proximity eases the difficulties of informational complexity. That is, terms become more accessible when counterparties are acquainted with (or are experts in) the relevant aspect of commerce where such terms are customary or standard.[78] However, tenuous relational proximity may make informational complexity more difficult. Where a party is a novice to a particular field of commerce, the higher the likelihood that terms peculiar to that field can impose a learning curve or information cost on that party.

Having described both parameters, we now shift to addressing how they may impact the communication of contract terms in the contemporary marketplace. As studies show, commercial contract documents have become more detailed and lengthier.[79] In some cases, they involve a combination of documents.[80] This may arise when a contract combines two or all three contracting methods. For example, one of the parties may dictate a set of terms. Some negotiated terms may be contained in a different document. Yet, another document may, by reference, seek to incorporate industry standards.[81]

Although informational complexity has become exacerbated in modern times, it has not escaped judicial notice as judges have expressed awareness of the lethargy for reading contract terms borne by actors in commercial settings.[82] However, it is fair to expect that, with strong relational proximity shared by parties, accessibility of terms increases. Hence, notice difficulties would be minimal.

Studies show that contract documents may be strategically drafted to capitalise on human biases to present terms in ways that may effectively

78 Cathy Hwang and Matthew Jennejohn, 'The New Research on Contractual Complexity' (2019) Capital Markets Law Journal 381.

79 Cathy Hwang and Matthew Jennejohn, 'Deal structure' (2018) 113; see also, Spencer Williams, 'Contracts as systems' (2021) Delaware Journal of Corporate Law 219.

80 Cathy Hwang, 'Unbundled bargains: multi-agreement dealmaking in complex mergers and acquisitions' (2016) 164 University of Pennsylvania Law Review 1403; see also *Alan Bates and Others Claimant v Post Office Ltd* [2019] EWHC 606 (QB), para 33.

81 Hwang and Jennejohn (n 78 above).

82 Chadwick LJ in *Lidl UK Gmbh v Hertford Foods Ltd & Another* [2001] EWCA Civ 938: 'as I suspect, common experience would suggest that busy executives often do not read the fine print in which standard conditions appear'; see also *Balmoral Group v Borealis* [2006] 2 Lloyd's Rep 629, para 339.

(and without illegitimate deception) lull offerees into contracts.[83] For example, terms may be presented in systematic ways that an otherwise shrewd party could possibly be led to underestimate risks or discount the likely implications of a term because an important term was presented as though it were unimportant. In negotiated contracts, lack of notice is not expected to be a persuasive argument a party raises. It is fair to assume that both parties were the joint authors of the contract's terms. If a party had seen a term whose value or implications they did not understand, they could have sought clarification. However, if the case is that a term was inserted into the contract contrary to the parties' joint agreement, then that party's remedy would be to a rectification of the contract to revise terms to which they did not consent. Therefore, it is fair to say that the signature rule suitably applies to negotiated contracts. Having addressed negotiated contracts, we now move to the other methods of contracting that concern us – ie industry standards and unilaterally dictated terms. The question is: how complicated could the communication of terms be when these contracting methods are deployed? To answer that question, we must go by *informational complexity* and *relational proximity* parameters.

Deducing guiding postulations

Regarding unilaterally dictated contracts, terms are dictated by the offeror to the offerees, and the offerees are simply term-takers. In such cases, there are real possibilities for the offerees to be confronted with informational complexity, and the tenuous transactional proximity between the parties likely amplifies this. As an observer noted concerning some unilaterally dictated terms deployed by certain banks in the United States (US) subprime mortgage market, 'some of the exotic new mortgages were so complicated that a person with a PhD in mathematics wouldn't understand them'.[84]

Industry standards stand in the middle of the extremes between negotiated and unilaterally dictated contracts. For this reason, the parameters of relational proximity and informational complexity would likely give us different results. This is because, in some cases,

83 See, for example, the statement of Salmon LJ in *Hollier v Rambler Motors* (1972) 2 WLR 401, 402: 'I do not think that defendants should be allowed to shelter behind language which might lull the customer into a false sense of security by letting him think – unless perhaps he happens to be a lawyer – that he would have redress against the person with whom he was dealing for any damage which he, the customer, might suffer by the negligence of that person.' See Kathleen C Engel and Patricia A McCoy, 'Turning a blind eye: Wall Street finance of predatory lending' (2007) 75 Fordham Law Review 2039, 2080; Hwang and Jennejohn (n 78 above).

84 Edmund Andrews, *Busted: Life Inside the Great Mortgage Meltdown* 1st edn (WW Norton & Company 2009) 77.

sophisticated parties with relatively equal bargaining positions may adopt industry standards that operate within close-knit or specialised markets. In such cases, the parties are very much akin to parties whose contracts are outcomes of negotiations. However, in other cases, offerees may have no choice but to accept industry standards because it is the governing standard in the industry or market in which an offeror operates. In the former case, the offerees are likely to be well-informed about the market dynamics and have the expertise and the resources to deal with the vagaries of the terms. In the latter case, the offerees may appear no different, in substance, from an offeree who is confronted with dictated terms. Yet, such offerees significantly differ from offerees (or addressees) of dictated terms. This is so because such offerees agreed to deal on industry standards, which were not authored by the offerors but are a product of market or industry consensus. Therefore, an offeror who deals with industry standards can be taken to have done everything reasonably necessary on their part to present the terms to the offeree.

We now move to test these general statements (derived from applying the parameters) against judicial outcomes concerning notice of terms. Towards that end, there are two lines of assessments to pursue. The first line addresses a central pillar of the signature rule: the assumption that businesspersons are generally sophisticated. The motive of this assessment is to critique that assumption by showing that sophistication is a gradation and is context-specific. If an offeree qualifies as possessing relevant transactional expertise or acquaintance, that would *generally* be sufficient to establish notice unless the offeree satisfactorily proves that the drafting of the terms was incomprehensible or difficult to understand (section four below deals with that discussion). The second line of the review relates to the requirement that a term must be onerous or unusual before it becomes a candidate for specific or emphatic notice to the offeree. As shall become apparent, this line is a fallback to the first (ie that of determining transactional expertise – where the offeree does not qualify as a transactional expert). This is because it relies on an inquiry based on the *informational complexity* parameter to ascertain whether relevant terms have been adequately communicated to the offeree. This discussion is pursued in section five below under the heading 'The requirement that terms must be onerous or unusual'.

THE ASSUMPTION THAT BUSINESSPERSONS ARE GENERALLY SOPHISTICATED

As noted, central to the signature rule's formalism is the view that contracting businesspersons can mostly be considered sophisticated. This judicial assumption of sophistication is untenable. Whether a businessperson is sophisticated should be assessed based on context, with regard had to the quality of transactional experience and knowledge attributable to the offeree concerning the bargain entered. As can be deciphered from relevant case law, courts often tend to get things right regarding sophistication in cases where relational proximity can be established. In such cases, if the offeree knows the offeror's market dynamics or pertinent peculiarities (or they ought to be known to the offeree), their sophistication can be taken as established. However, where there is no such evidence of relational proximity, then assumptions of sophistication cannot hold.

This section addresses how an appropriate level of sophistication or transactional expertise may convincingly be established and used to cement sufficient notice. Cases dealing with the ascertainment of sophistication can be divided into two: those concerning industry standards and those dealing with unilateral dictations. We shall deal with each in turn. The analysis draws on judicial decisions relating to both signed and unsigned contracts (or terms) to justify the postulations made in this section. This is because the selected cases help illustrate the informational dynamics pertinent to industrial standards and unilaterally dictated terms.

Industry standards and offeree sophistication

Given the widespread use of industrial standards in a specific industry or market, their adoption may sometimes extend beyond national application and encompass transnational application.[85] As Braithwaite describes it, 'standard form contracts may be understood as a set of binding norms that are generated privately'.[86] Their industry or market popularity reflects the need to expect their audiences to be experts in the relevant commercial field and its dynamics.[87] Therefore,

85 See, Briggs J's statement concerning ISDA standard regime in *Lomas v JFB Firth Rixson* [2011] 2 BCLC 120, [53].

86 Joanne Braithwaite, 'Standard form contracts as transnational law: evidence from the derivatives markets' (2012) 75 Modern Law Review 779, 780.

87 See *AIB Group (UK) Ltd v Martin* [2002] 1 WLR 94, 96 per Lord Millett: 'A standard form is designed for use in a wide variety of different circumstances. It is not context-specific. Its value would be much diminished if it could not be relied upon as having the same meaning on all occasions. Accordingly the relevance of the factual background of a particular case to its interpretation is necessarily limited.'

they are assumed to possess knowledge of terms contained in specific sets of standard terms. For that reason, it is fair to think that when an offeror makes a reference that a designated or specified set of industry standards governs, the offeree who contracts with the offeror must be taken to have sufficient notice of all the provisions contained in the said set of terms. This is because it is fair to assume that persons agreeing to deal with industry standards are sophisticated, even if they are not necessarily so.

The main justification for the assumption of offeree sophistication in this regard is that the terms are not a product of the offeror's authorship. For this reason, offerors are not required to do anything further to present the terms beyond intimating to the offeree the standard's adoption. In turn, the offeree can be assumed to have had due notice of the terms and their contents by agreeing to them.

In *Photolibrary Group Ltd v Burda Senator Verlag GmbH*,[88] the claimant, suppliers of photographic transparencies, had agreed to deal with the defendants (the offerees) using industry standards recommended by the British Association of Picture Libraries and Agencies (BAPLA). One of the terms of the BAPLA standard catered to the compensation of the offeror for transparencies lost by offerees, and the claimant sought to be compensated based on that term. The offerees contested the term's validity, asserting that it had not been adequately brought to their notice and, as such, not incorporated. The offerees' argument of lack of notice was rejected because they understood the offeror dealt based on the BAPLA terms. Therefore, the offerees 'must be taken to have instructed the obtaining of the transparencies on those terms'.[89] The English Court of Appeal came to a similar position in *Circle Freight International Ltd v Medeast Gulf Exports Ltd*[90] by rejecting the offeree's argument that an exclusion clause contained in the relevant standard terms should not apply as specific notice of the said clause had not been provided. The court ruled that the offeree was bound by the clause given their awareness that the standard terms in issue governed in the relevant industry and because the offeror had referred to them as regulating the bargain between the parties.

Sometimes, an offeree may not be aware or conversant with specific terms in industry standards. This may arise because the offeree is a novice in the said industry. Concerning such a situation, the reasoning expressed by Mellish LJ in *Parker v South Eastern Rail Co*[91] is pertinent. In that case, the judge cited the analogy of a novice in the shipping industry who is presented with a bill of lading and ignorant of

88 [2008] 2 All ER (Comm) 881.
89 Ibid 896.
90 [1988] 2 Lloyd's Rep 427.
91 [1874–80] All ER Rep 166.

its purpose and terms. Concerning such a person, the Lord Justice said 'the shipbroker or the master delivering the bill of lading is entitled to assume that the person shipping goods has that knowledge'.[92] Further, that 'such a person must bear the consequences of his own exceptional ignorance, it being plainly impossible that business could be carried on if every person who delivers a bill of lading had to stop to explain what a bill of lading was'.[93]

Dictated terms and offeree sophistication

In unilateral contracts, an offeree's expertise may be established based on the strength of knowledge attributable to the offeree concerning the offeror's terms. A prime indicator of sophistication or expertise in such a case is the combination of the facts that the offeree has contractual terms identical to those of the offeror and that the offeree deals in the same market as the offeror. An excellent example of this is seen in *Allen Fabrications v ASD*.[94] In that case, the English Court of Appeal reasoned (among other things) that since the offeree had operated in the same market as the offeror and had exclusion clauses like that of the offeror, there was a sufficient basis for the offeree to be on notice of the offeror's exclusion clause.[95] We find similar outcomes in *Watford Electronics Ltd v Sanderson CFL*[96] and *Balmoral Group v Borealis*.[97]

Another possible indicator of offeree sophistication or expertise in a relevant market is that the offeree has had an established history of dealings with the offeror or other persons in the same market.[98] But an inconsistent or infrequent history of dealings is not representative of expertise. In *Carlsberg-Tetley Brewing Ltd v Gilbarco Ltd*,[99] the court refused the offeree's claim that they did not have notice of the offeror's exclusion clause. The court did so on the ground that the knowledge and expertise possessed by the offeree 'in engineering matters concerning underground pipes feeding fuel to fuel pumps was greater than that of the defendants'.[100] Also, the offerees 'had Chartered Engineers on their staff and had themselves originally designed the system of tanks and pipes'.

---

92 Ibid 169.
93 Ibid.
94 [2012] EWHC 2213 (TCC).
95 Ibid paras 64, 76.
96 [2001] EWCA CIV 317, para 48.
97 *Balmoral Group v Borealis* (n 82 above) para 147.
98 *SIAT v Tradax* [1978] 2 Lloyd's Rep 470.
99 [1999] 3 WLUK 593 (No 1998 TCC No 445).
100 Ibid para 25.

Similarly, in *British Crane Hire Corp v Ipswich Plant Hire*,[101] a case of an orally formed contract, Lord Denning took the factors of market expertise and a history of dealing with other parties in the market as sufficient in finding notice. Most pertinently, his Lordship said:

> But here the parties were both in the trade and were of equal bargaining power. Each was a firm of plant hirers who hired out plant. The defendants themselves knew that firms in the plant-hiring trade always imposed conditions in regard to the hiring of plant: and that their conditions were on much the same lines.[102]

We find a similar outcome in *HIH v New Hampshire*.[103] In that case, the subject matter in dispute was one of reinsurance. In their unilateral terms, the offeror included an exclusion clause customary in the London reinsurance market. As the court described it, the London market 'was more diverse, and included non-investment grade transactions, such as the instant insurances, where there was real risk and a higher rate of premium'.[104] For these reasons, players in the London market were more comfortable using exclusion clauses. However, the offeree was more conversant with practices in the US market, which, as the court described it, were 'concerned with investment grade security transactions, the relevant insurers were risk-averse, the premium rates were low, but the sums insured very high'.[105] In the US market, exclusion clauses were considered unusual. The offeree denied notice of the exclusion clause in issue as they had not specifically been informed concerning them. But the court rejected that argument, ruling that the term, while not standard or customary in the said market, was 'by no means unknown but something that the market would recognise'.[106] Given the offeree's general expertise in the reinsurance market, the court expected circumspection from the offeree.

On the other hand, where the offeree is not a market expert, it would be improper to attribute awareness of terms to the offeree. This is a situation, such as that which Sir Eric Sachs described as being one in which the offeror and the offeree are 'in wholly different walks of life'.[107] That is, 'where one, for instance, is an expert in a line of business and the other is not'.[108] In such a case, it is expected that notice of peculiar terms must be brought to the offeree. A commercially unsophisticated

101 [1974] 2 WLR 856.
102 Ibid 861.
103 [2001] EWCA Civ 735.
104 Ibid para 200.
105 Ibid.
106 Ibid para 214.
107 *British Crane Hire Corp v Ipswich Plant Hire* [1974] 2 WLR 856, 863.
108 Ibid.

offeree may not be deemed fixed with knowledge of peculiar terms, even if such terms are prevalent in the relevant market.[109] It is fair and reasonable for such an unacquainted party to assume or expect each offeror to have different terms. This allows such an offeree to compare the terms of various offerors before transacting. Therefore, an offeror may not argue that an offeree's history of dealing (with other offerors) in the relevant market is sufficient to fix that offeree with notice of otherwise peculiar (or essential) terms that the offeror adopts but which have not been effectively brought to the offeree's notice.[110]

To ensure adequate notice, the offeror must do all that is reasonably expected to be done in the context of bringing notice of the terms to the offeree. Similarly, an offeree who has dealt with a particular offeror numerous times but was never specifically informed of peculiar terms cannot reasonably be considered to be fixed with notice of such terms because of the history of dealings with that offeror. For notice of such peculiar terms to be effective, the offeror must ensure effective awareness of them, at least once, throughout their dealings with the offeree. This position finds corroboration in the views expressed by Lord Devlin in *McCutcheon v David Macbrayne*[111] and Lord Guest in *Hardwick Game Farm v Suffolk Agricultural Poultry Producers Association*[112] that: 'Previous dealings are relevant only if they prove knowledge of the terms, actual and not constructive, and assent to them.'

In essence, while a set of unilateral terms previously used in an established course of dealing may be incorporated, peculiar salient clauses may not be considered incorporated and enforceable if they are not specifically made known to the offeree or if they are expressed using insufficiently clear wordings. From the preceding analysis, we may deduce two guiding principles. One is that offeree sophistication cannot be assumed when the offeree is not an expert in the market to which a bargain relates. The second is that, even if an offeree is an expert, the fact that the wordings of terms are elusive or tricky may destroy the value of relational proximity (ie expertise) in establishing notice.[113]

Submissions

The central plank of this section is that relational proximity can be taken as alleviating informational complexity. Also, counterparties' market expertise is a viable indicator of relational proximity. The

109 *Apps v Grouse Mountain Resorts Ltd* 2020 BCCA 78, para 70–84.
110 *Blu-Sky Solutions* (n 2 above) para 109(iii).
111 [1964] 1 WLR 125 (UK HL).
112 [1968] 3 WLR 110, 154.
113 *Thompson v London, Midland and Scottish Railway Co* [1930] 1 KB 41, 52.

question is: what is the yardstick by which we may identify a party as a market expert? For our purposes, one may describe an expert in a market (or industry) as either a buyer or seller of a subject matter (ie goods or specific service) that is integral to the operation of their trade and who engages with the said subject matter with reasonable scale and frequency.[114] The fact that an addressee of unilateral terms does not have equal bargaining powers with the term-giver should not negate a finding that the addressee is an expert.

Notwithstanding such possible gaps, all that is material is that the addressee is a commercial undertaking.[115] Also, for reasons already explained, a 'market novice' who enters a contract governed by industry standards will be considered an expert for that transaction. Where an addressee qualifies as an expert, it is fair to attribute to that party notice of terms that they ought reasonably to expect as customary in the said market. It is submitted that the burden of proving a lack of commercial expertise should be placed on an addressee to discharge.

This submission corresponds with the postulations of prominent scholars who deploy socio-economic analysis in contract law scholarship. Two relevant features are often identified by them – learning and network effects.[116] Learning effects relate to the frequency with which a species of terms is used in a market/industry. On the other hand, network effects relate to expectations borne by market/industry participants that that species of provisions would be used in prospective contracts or the general future. In other words, learning effects imbue or nurture network effects. The upshot is that where a class of terms is commonplace in a market, a counterparty would struggle to deny notice or anticipation of its usage. The only circumstance where such a counterparty may successfully deny notice is where the terms are unusual or inscrutably drafted. Such a situation would represent a departure from the terms' learning effects and, thus, lack network effects.

---

114 *Lynch Roofing Systems Ltd v Bennett & Son Ltd* [1999] 2 IR 450; *Scheps v Fine Art Logistic Ltd* [2007] EWHC 541 (QB); C*apes (Hatherden) Ltd v Western Arable Services Ltd* [2009] EWHC 3065 (QB).

115 *McCrory Scaffolding Ltd v McInerney Construction Ltd* [2004] IEHC 346.

116 See Kahan and Gulati (n 66 above) 1127–1133; Michael Klausner, 'Corporations, corporate law, and networks of contracts' (1995) 81 Virginia Law Review 757.

THE REQUIREMENT THAT TERMS MUST BE ONEROUS OR UNUSUAL

We now shift to those situations where the relevant quality of transactional expertise or sophistication required of the offeree is not proven. For that reason, the inquiry would then depend on the parameter of *informational complexity* alone. Here, the pivotal inquiry would be: 'whether notice of relevant terms has been effectively communicated to the offeree'. This section also addresses cases where industry standard terms and dictated terms bear incomprehensible or elusive drafting that the average 'transactional expert' would have difficulty understanding.

Where a party qualifies as possessing transactional expertise (following the analysis and postulations of section four above), the need for bringing peculiar terms to the notice of that party is dispensed with. Some transactional contexts exemplify this view. A prime example would be investment arrangements. In such dealings, it is only fair to take it as the offeree's responsibility to ensure they understand what they are getting into. In other words, barring fraud or misrepresentation on the part of offerors (in such cases), offerees must ensure that they clearly understand the rules of such arrangements. A significant explanation for such a position is that such arrangements present offerees with opportunities for an economic windfall.[117] The zero-sum nature of such agreements was described in the Australian High Court in *Kakavas v Crown Melbourne Ltd*[118] as a kind where each party's unmistakable purpose is to 'inflict loss upon the other party to the transaction'.

Similarly, in the Canadian case of *Chen v TD Waterhouse*,[119] where the offeree had argued that certain aspects of an investment arrangement had not been brought to their notice, the court rejected the argument for specific notice. The court said, among other things, that the transaction involves 'substantial sums of money for risky and highly regulated stock market trading' and, as such, 'is not at all a "hurried, informal" affair analogous to renting a car at a pick-up counter'.[120] Therefore, it was incumbent upon the offeree to have spent time and effort understanding the documents connected with the arrangement.

---

117 Mark Gergen, 'A defense of judicial reconstruction of contracts' (1995) 71 Indiana Law Journal 45, 47.

118 [2013] HCA 25.

119 [2020] OJ No 1037.

120 Ibid para 29.

As can be gleaned from Rudanko's exposition on digitised contracts, particularly high-frequency trading, they are much analogous to gaming contracts.[121] As he explains:

> In the automated trading in securities markets, offers and acceptances are fed in the system that then matches the corresponding inputs into contracts, which are in the next stage settled in an automated netting system, distributing the net assets to the participants in trading.

He goes further to say: 'In the trading system, no traditional contract relations between identifiable parties can be detected, but there is no doubt about the existence of real sale of securities transactions.' For this reason, it appears convenient to refer to cases on gaming contracts where the notice of contested terms was in issue. A good example is the English gaming case of *O'Brien v MGN Ltd*[122] where the court reasoned that there was no need for the offeror (game operator) to bring the contested terms to the notice of the offeree (player). The player, the court reasoned, could (and should) have discovered the rules themselves, as all gaming arrangements operate based on specific rules.

But as reasoned in *Green v Petfre (Gibraltar) Ltd*,[123] a term-taker cannot reasonably be said to have notice of terms in a gaming agreement where such terms are patently 'obscure and unclear'. In some recently decided English cases, the courts ruled that the operator (or their agent) of an investment scheme owed no obligation to investors to advise them on the risks they are likely to encounter in participating in the scheme.[124] However, a duty known as a 'mezzanine duty to advise' only arises when the operator voluntarily assumes the responsibility of explaining the scheme's risks to the offeree.[125] Only in such a situation of the voluntary assumption must the operator be fully transparent in disclosing the relevant risks to the offeree.

There are other cases where heightened notice expectations are not placed on offerors. These are cases where any prudent and reasonable observer would be satisfied that the offeror had done everything that was objectively sufficient to bring all the terms, including

121 Matti Rudanko, 'Smart contracts and traditional contracts: views of contract law' in Marcelo Compagnucci, Mark Fenwick and Stefan Wrbka (eds), *Smart Contracts: Technological, Business and Legal Perspectives* (Hart 2021) 59, 67.

122 [2001] EWCA Civ 1279.

123 [2021] EWHC 842 (QB); see also, Joanne McCunn, 'Incorporation of terms: time for Blu-Sky thinking?' (2022) Lloyd's Maritime and Commercial Law Quarterly 189.

124 *Crestsign Ltd v National Westminster Bank & Royal Bank of Scotland* [2015] 2 All ER (Comm) 133; *Thornbridge Ltd v Barclays Bank plc* [2015] EWHC 3430 (QB).

125 *Property Alliance Group Ltd v Royal Bank of Scotland* [2016] EWHC 3342 (Ch).

salient ones, to the notice of the offeree. For example, where the terms are thematically arranged, highly readable and expressed in a few paragraphs.[126] Where there is no time pressure on the offeree, enabling them to carefully read the contract terms and ask questions about terms not understood before signing.[127] Also, sufficient notice would be deemed as served where salient terms were expressed in an 'almost apocalyptic' fashion.[128]

However, in cases where transactional expertise cannot fairly be attributed to the offeree, then it is incumbent upon the offeror to ensure sufficient notice of salient terms to the offeree. Notice is contestable in cases where salient terms that are vital in shaping the incentives of (non-expert) offerees to accept or refuse a contractual offer are not specifically brought to their attention. According to prevailing views, only onerous or unusual terms are required to be brought to the notice of the offeree in situations of contestable notice. It complicates matters that there is no discernible judicial consensus on what amounts to 'onerous' or 'unusual' terms.[129] Judges have defined those qualities differently. Notwithstanding the lack of uniformity in definition, it is observable that courts expect terms that qualify as such to be of a nature that is (totally) unexpected, commercially perverse, or something close to outlandish.[130] For this reason, remedial clauses, exclusion and exemption clauses, arbitration clauses, and commission-withholding clauses have been judged as not being onerous and, thus, not requiring specific notice.[131]

Judges are aware of the empirical reality that people do not pay rapt attention to contractual terms but instead focus on ones that appear salient, so the rules of notice should be reformed with this reality in mind. Thus, in those cases where offerees are non-experts or otherwise possess expertise (but the terms are presented using tricky or inscrutable wordings), specific notice must be required for

126 *Bedford Investments Ltd* (n 12 above) para 94; *Carlsberg-Tetley Brewing Ltd v Gilbarco Ltd* [1999] EWHC J0322-5; *Rogers Cable Communications Inc v York Condominium Corp* [2005] OJ No 4099.

127 *Canadian Linen and Uniforms v Taurus* [2020] AJ No 497; see also, *Burkshire Holdings Inc v Ngadi* [2021] OJ No 2781.

128 See *BA Kitchen Components Ltd v Jowat (UK) Ltd* [2021] NIQB 3, para 23; see also *Goodlife Foods Ltd v Hall Fire Protection Ltd* [2018] EWCA Civ 1371, para 53.

129 See *Streford v Football Association* [2007] EWCA Civ 238; *Trebor Bassett Holdings Ltd and Another v ADT Fire and Security plc* [2011] EWHC 1936 (TCC).

130 *O'Brien v MGN Ltd* [2001] EWCA Civ 1279; see also *Trebor Bassett Holdings* (n 129 above).

131 See *Peninsula Business Services Ltd v Sweeney* [2004] IRLR 49; *Streford v Football Association* (n 129 above); see also *Gregg & Co (Knottingley) Ltd v Emhart Glass Ltd* [2005] EWHC 804 (TCC).

salient terms. That is, terms that would have been vital to the offeree's exercise of choice between competing offerors, particularly those likely to impact the allocation of benefits and burdens between offerees and offerors.[132] These are terms likely to have material effects on the quality of contractual rights and obligations of an offeree. They include but are not limited to, all remedial clauses,[133] exclusion clauses,[134] termination clauses, arbitration clauses[135] and restrictive covenants.[136] This position is akin to Spencer's.[137]

As can be deduced from Eisenberg, contractual addressees are unlikely to treat clauses relating to consequences of non-performance or default as necessary to their decisions to enter a contract.[138] This is because people often discount the possibility of breach or default. For this reason, offerees will likely treat remedial clauses as non-salient because they consider a breach a low-probability event. Therefore, an offeror who intends to incorporate such a term must bring it to the notice of the offeree in a way that emphasises its salience. Thus, in *Bridge v Campbell Discount*,[139] Lord Denning categorically expressed an identical position. He reasoned that, although the offeror (seller) and the offeree (hirer) had effectively entered a contract, the offeree had not consented to the remedial clause inserted into the contract.[140] This is because, as Lord Denning rightly reasoned, the remedial clause was remote to the contractual considerations of the offeree when entering the contract. In effect, the offeree had only consented to or was mindful of the salient express terms, such as the price and dates at which payment was due. We also find a similar position expressed by Lord Denning in *United Dominions Trust (Commercial) Ltd v Ennis*.[141] For such terms to be salient in the calculations of the offeree, it should be expected that the offeror gives adequate notice of such terms.

---

132 Russell Korobkin, 'Bounded rationality, standard form contracts, and unconscionability' (2003) 70 University of Chicago Law Review 1203, 1206.

133 *Madison Homes Cornell Rouge Ltd v Chi-Hong Stanley Ng* [2021] OJ No 2369; *MacQuarie Equipment Finance (Canada) Ltd v 2326695 Ontario Ltd* [2020] OJ No 720.

134 *Luminary Holding Corp v Fyfe* 2021 BCSC 167; *Robophone Facilities Ltd v Blank* [1966] 1 WLR 1428, 1446.

135 *Kaye v Nu Skin UK Ltd* [2009] EWHC 3509 (Ch).

136 *SI Systems Partnership v Geng* [2020] OJ No 5744.

137 J R Spencer, 'Signature, consent, and the rule in L'Estrange v Graucob' (1973) 32 Cambridge Law Journal 104.

138 Melvin Eisenberg, 'The limits of cognition and the limits of contract' (1995) 47 Stanford Law Journal 211, 243.

139 [1962] 2 WLR 439.

140 Ibid 446, 457.

141 [1967] 3 WLR 1, 7.

Submissions

The import of this section is that where an addressee proves to be a market 'novice' (not being an expert), then it would be incumbent on the term-giver to demonstrate that they had done everything reasonably possible to put a prudent person in the situation and circumstances of the addressee on notice of terms that are otherwise salient or material.[142] The argument is not that an offeror must ensure that the addressee reads the terms. Instead, the offeror has objectively done everything practically possible to ensure adequate notice to the addressee.

One major factor is vital here, and that factor is effective informational presentation. This factor is vital both in digital settings and paper-based ones.[143] In the face of informational complexity, how material terms are presented to addressees would impact how effectively they discern their material rights and obligations. This would then shape their ability to exercise their choice (or arrange their affairs) at the pre-contractual and post-formation stages. The assessment of what qualifies as a practical presentation of terms would depend on the facts of each case. In digital settings, the addressee may use a combination of features to improve the presentation of salient terms. The transactional interface may use colour codes or schemes to distinguish terms based on their materiality. They may use pings and images.[144] They may also use navigational aids and input controls (eg checkboxes requiring an offeree's confirmation of material terms before they can proceed to pay or complete contractual formation). In paper-based settings, the brevity of the statements, thematic arrangements of terms, the use of colour codes and distinct fonts and so on would matter for an effective informational presentation. Thus, the burden of proving that reasonable measures were taken to put a prudent person in the addressee's condition and circumstances on notice is to be shouldered by the term-giver.

Deductions from the forgoing analyses on the interactions between relational proximity and informational complexity or accessibility may be represented in the Figure 1.

142 See the US cases of: *Scotti v Tough Mudder Inc* 63 Misc 3d 843 (2019); B*erkson v GOGO LLC*, 97 F Supp 3d 359 (2015).

143 Nancy Kim, 'Digital contracts' (2019) 75 The Business Lawyer 1683.

144 See Woodrow Hartzog, 'Website design as contract' (2011) 60 American University Law Review 1635, 1667.

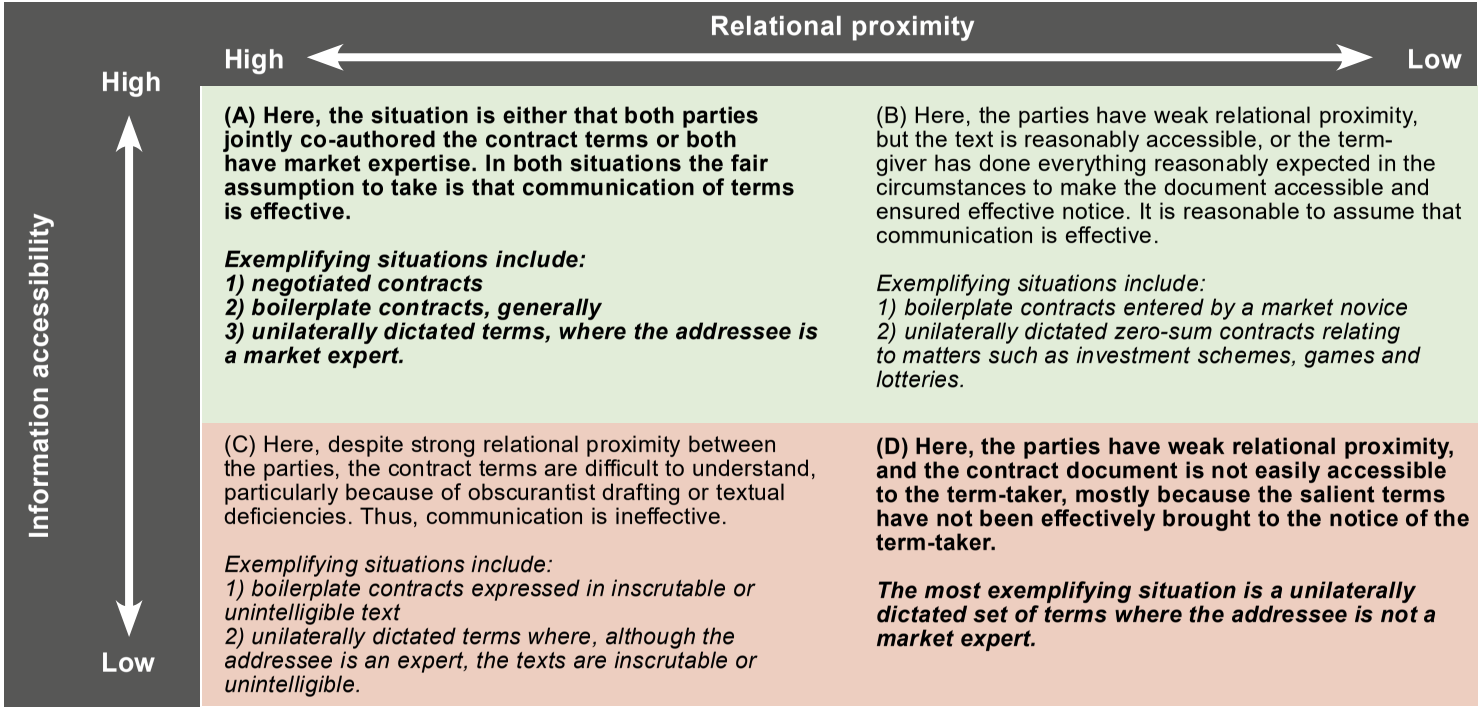

Figure 1: (A) and (D) capture standard cases. In most cases where relational proximity is strong, as in (A), informational accessibility follows. Where relational proximity is weak, as in (D), it often follows that informational accessibility is difficult. Relatively common exceptions are outlined in (B) and (C).

RECOMMENDATIONS AND CONCLUSION

From the foregoing analysis, it can be discerned that the signature rule takes an oversimplified view regarding the effective communication of contract terms. In that regard, the following are postulated:

1. That signature should generally be conclusive of consent in cases of negotiated and industry-standard terms. In other words, there should not be an inquiry into the sufficiency of notice in those cases.
2. In cases of industry standards, there may be the need to inquire into the sufficiency of notice where it can be proved that the wordings of the particular contract terms are so abstruse or complex as to defy the understanding of the average expert in the market or industry.
3. In cases of unilateral terms, offerees should be deemed sufficiently notified of all terms, including peculiar ones, in situations where it can demonstrably be shown that they are conversant with market dynamics and have had a series of dealings with the term-givers/offerors or alternative offerors with similar terms. It should also be probative of notice if offerees are market experts with their terms like the offeror's.
4. Regarding unilateral terms, where the offerees are not market experts, then terms that would otherwise have been salient to their consideration in entering the contract should be specifically brought to their awareness for notice to be adequate. In other words, the exacting qualities of 'onerous' and 'unusual' should not be the touchstone for determining the requirement of specific disclosure of terms to offerees.

With a context-specific test such as that proposed here, we can prevent using contract terms as economic traps for unwary offerees. Most importantly, in unilateral terms where offerees are not market experts and those of industry standards expressed in obscure ways, there should be heightened disclosure expectations for notice to be deemed sufficient. The rationale for this submission is that non-price terms are often equivalent or complementary to price terms. Thus, non-price terms contained in unilateral and industry-standard terms may facilitate the extraction of unearned economic rent from offerees to offerors, primarily when such terms are expressed as though there were unimportant terms. This endows such terms with potential ambush effects. The real possibility of offerees discounting the cost of such terms is apt to equip offerors with market power to increase market traction through their strategic presentation of contract terms in a commercially attractive fashion.

Northern Ireland
Legal Quarterly
Autumn Vol. 74.3 (2023) 472–509
Articles DOI: 10.53386/nilq.v74i3.1018

Transitional justice from above and below: exploring the potential glocalising role of non-governmental organisations through a Northern Ireland case study*

Fiona McGaughey
University of Western Australia
Rachel Rafferty
University of Derby
Amy Maguire
University of Newcastle, Australia
Correspondence email: fiona.mcgaughey@uwa.edu.au

ABSTRACT

Non-governmental organisations (NGOs) have the potential to play a unique local-to-global role – 'glocalisation' – as intermediaries to international human rights bodies. NGOs are also significant actors in transitional justice societies and have the potential to make a positive contribution by linking local and global dimensions of transitional justice. Moving beyond analysis of legal transitional justice responses through domestic and regional courts and the United Nations (UN) treaty body individual complaints, this article calls for consideration of the role NGOs can play in transitional justice when they connect to quasi-judicial and political UN human rights bodies through state reporting mechanisms. As the work of these bodies is quite reliant on NGO expertise, particularly local NGOs, the article examines Northern Ireland NGO engagement with UN state reporting mechanisms regarding transitional justice.

The article concludes that, while the distinction between the state-centric review at the UN level and grassroots activities at a local level is ultimately a false dichotomy, this divide seems to be operating in practice in the case of Northern Ireland. We make a distinction between rights-focused and reconciliation-focused NGOs and find that reconciliation-focused NGOs in particular are largely absent from international reporting frameworks. We argue that NGOs have the potential to play a unique local-to-global role, 'glocalisation', but this only works if local NGOs are enabled and encouraged to engage at a global level. Hence, we recommend that, where possible, local NGOs must be involved in both grassroots activities and international monitoring via the UN in order to exploit their glocalising potential.

Keywords: transitional justice; NGOs; United Nations; human rights; UN Human Rights Council; peace; Northern Ireland; treaty bodies.

* We acknowledge the careful assistance of Jessica Heaney in the development of this article. We also thank the reviewers for their helpful comments.

INTRODUCTION

Transitional justice aims to deal with the past in the aftermath of violent conflict or dictatorial regimes and, given the diversity of strategies employed globally, is now understood as an umbrella term for a range of approaches.[1] Contemporary critique also recognises that, whereas victor's justice and the role of amnesties were initial focuses, this is now supplemented by other concerns such as the underlying politics, economic justice and the need to balance local and international agency.[2] Increasingly, it is recognised that an effective transitional process must engage with society as a whole, going beyond purely legal measures to promote a shift in the relationship between citizens and the state and between different social groups. Such socio-legal conceptualisations of transitional justice overlap considerably with the theory and practice of peacebuilding, a field that aims to support the achievement of sustainable peace and reconciliation in post-conflict societies.[3] Similar to the expansion of transitional justice beyond purely legal concerns, peacebuilding is increasingly recognised as requiring transformative change in social relations as well as institutions.[4] Scholarship in the fields of both transitional justice and peacebuilding debates the relative merits of either a 'top-down' or 'bottom-up' approach to achieving desired social changes.[5]

The dominant discourse is one of dichotomy; that transitional justice and related peacebuilding activities are best approached either from below – including efforts by civil society, grassroots organisations and communities – or above – including the activities of governments, funders and international organisations.[6] However, this distinction is too simplistic and, increasingly, the capacity of civil society to not only

---

1 Susanne Buckley-Zistel, Teresa Koloma Beck, Christian Braun and Friederike Mieth, 'Transitional justice theories: an introduction' in Susanne Buckley-Zistel et al (eds), *Transitional Justice Theories* (Routledge 2013) 1.

2 Khanyisela Moyo, *Postcolonial Transitional Justice: Zimbabwe and Beyond* (Routledge 2019).

3 Dustin Sharp, 'Beyond the post-conflict checklist: linking peacebuilding and transitional justice through the lens of critique' (2013) 14 Chicago Journal of International Law 165.

4 Brandon Hamber and Elizabeth Gallagher (eds), *Psychosocial Perspectives on Peacebuilding* (Springer 2014).

5 See, for example, Nevin Aiken, Identity, *Reconciliation and Transitional Justice: Overcoming Intractability in Divided Societies* (Routledge 2013); Kieran McEvoy and Lorna McGregor (eds), *Transitional Justice From Below: Grassroots Activism and the Struggle for Change* (Hart 2008).

6 See generally the scholarship that aims to elevate the voice of communities 'on the ground', as a response to formal legal institutions and international mechanisms, such as McEvoy and McGregor (n 5 above).

challenge and critique top-down legalistic approaches to transitional justice, but also complement and expand on them, is acknowledged.[7] Similarly, the concept of 'hybrid peace', whereby internationally supported peace operations and local approaches to peace interface, has also gained traction in debates on local ownership in peacebuilding.[8]

The definition of transitional justice used here is from the then United Nations (UN) Secretary General Kofi Annan, who in 2004 defined it as comprising the 'full range of processes and mechanisms associated with a society's attempts to come to terms with a legacy of large-scale past abuses, in order to ensure accountability, serve justice and achieve reconciliation'.[9] The Secretary General acknowledged that these 'may include both judicial and non-judicial mechanisms, with differing levels of international involvement (or none at all)'.[10] Despite this broad definition of transitional justice, some scholarship focuses solely on legal responses, particularly prosecutions, without acknowledging that this is only one (optional) part in a larger process.[11] Increasingly, however, transitional justice is recognised as a process that aims to not only (re)establish legal protections for human rights, but also to achieve a shift in attitudes and relationships among citizens towards support for peaceful coexistence, resulting in social reconciliation.

As can be seen from these definitions, transitional justice shares a central goal with the field of peacebuilding which also aims to achieve reconciliation. Transitional justice is increasingly seen as one of a number of peacebuilding tools that can assist post-conflict societies, alongside other initiatives such as demobilisation, disarmament and reintegration (DDR) or security sector reform (SSR).[12] However, the fields have developed largely in isolation from one another, and scholars have identified clear differences in how each field aims to achieve reconciliation. While transitional justice focuses on accountability and legal justice for victims, peacebuilding places greater emphasis on achieving political cooperation, reintegrating perpetrators and

---

7 Kieran McEvoy, 'Commentary on locality and legitimacy' in Nicola Palmer, Phil Clark and Danielle Granville (eds), *Critical Perspectives in Transitional Justice* (Intersentia 2012) 311.

8 Roger Mac Ginty, 'Hybrid peace: the interaction between top-down and bottom-up peace' (2010) 41(4) Security Dialogue 391.

9 United Nations Security Council, 'The rule of law and transitional justice in conflict and post-conflict societies: report of the Secretary-General' UN Doc S/2004/616, 23 August 2004.

10 Ibid.

11 See, for example, Jaya Ramji-Nogales, 'Designing bespoke transitional justice: a pluralist process approach' (2010) 32(1) Michigan Journal of International Law 3.

12 Sharp (n 3 above).

building trust between communities.[13] This is reflected in differences in international support, where progress in human rights and transitional justice is monitored through UN reporting frameworks while peacebuilding work receives less formalised support through the UN peacebuilding commission – an intergovernmental advisory body that supports peace efforts in conflict-affected countries.[14]

Opportunities for greater connection between these two fields in practice may be being missed currently. Non-governmental organisations (NGOs) that view their primary remit as peacebuilding may not perceive their potential to contribute to, and benefit from, UN human rights reporting mechanisms that support transitional justice and human rights.[15]

A distinction has been identified between NGOs which are predominantly *rights-based* in their approach to peacebuilding and those which favour a *reconciliation-based approach*.[16]

Rights-based NGOs (rights NGOs) naturally tend to align their work with a legalistic approach to transitional justice and may engage little with wider aspects of peacebuilding. For example, they have pursued strategic litigation in domestic or regional courts, or in international law by lodging petitions with UN treaty bodies. With regard to the conflict in Northern Ireland, much legal practice and scholarship has been dedicated to the impact of the domestic legal system on the conflict, the use of the European Court of Human Rights and, to a lesser extent, petitions to UN treaty bodies.[17]

Such legal avenues are important in the suite of transitional justice tools, but the use of international human rights law and

13 Catherine Baker and Jelena Obradovic-Wochnik, 'Mapping the nexus of transitional justice and peacebuilding' (2016) 10(3) Journal of Intervention and Statebuilding 281.

14 Simultaneously established by resolutions of the General Assembly and Security Council: United Nations General Assembly (UN GA) (2005) Resolution adopted by the General Assembly on 20 December 2005 (A/RES/60/180) and United Nations Security Council (2005) Resolution 1645 (S/RES/1645).

15 Maggie Beirne and Colin Knox, 'Reconciliation and human rights in Northern Ireland: a false dichotomy?' (2014) 6(1) Journal of Human Rights Practice 26.

16 Ibid.

17 See, for example, Charles Carlton, 'Judging without consensus: the Diplock Courts in Northern Ireland' (1981) 3(2) Law and Policy 225; Antoine Buyse and Michael Hamilton (eds), *Transitional Jurisprudence and the ECHR: Justice, Politics and Rights* (Cambridge University Press 2011); Michael O'Boyle, 'Torture and emergency powers under the European Convention on Human Rights: Ireland v the United Kingdom' (1977) 71(4) American Journal of International Law 674; Onder Bakircioglu and Brice Dickson, 'The European Convention in conflicted societies: the experience of Northern Ireland and Turkey' (2017) 66(2) International and Comparative Law Quarterly 263.

international human rights bodies to lobby for transitional justice in a non-litigious way, and the specific potential of reconciliation-based NGOs (reconciliation NGOs) to contribute, has typically received less attention. Reconciliation NGOs are more likely to engage in peacebuilding activities that focus on building positive relationships between communities and are less likely to engage with the legalistic aspects of transitional justice.[18]

However, this should not mean that international human rights law cannot be leveraged by reconciliation NGOs. In particular, the quasi-judicial role of UN human rights treaty bodies and the political role of states on the UN Human Rights Council are fora in which reconciliation NGOs could make a significant and unique contribution due to their connection to local communities and their insight into how legal decisions impact on intercommunal relationships at local level.

It is the potential of NGOs to play a unique role in connecting top-down legalistic transitional justice initiatives and bottom-up reconciliation efforts that we explore in this article. The term 'glocalisation'[19] has been adopted to express the intermediary role that NGOs can play between the global and local. At the global level, UN bodies are germane to the promotion and protection of human rights internationally. As such, post-conflict societies, where widespread human rights abuses have taken place, and where human rights may still be compromised, should warrant both scrutiny and support. The key monitoring mechanism of UN human rights bodies is that of state reporting, whereby a state's compliance with its international obligations is monitored by reviewing periodic reports submitted by

18 Beirne and Knox (n 15 above).

19 Initially a Japanese business term meaning the creation of products or services intended for the global market, glocalisation was later adopted by sociologists. See Habibul Haque Khondker, 'Glocalization as globalization: evolution of a sociological concept' (2004) 1(2) Bangladesh e-Journal of Sociology 14. The Oxford English dictionary definition of glocalisation is: 'The action, process, or fact of making something both global and local; specifically the adaptation of global influences or business strategies in accordance with local conditions; global localization': 'Glocalisation' (*Oxford English Dictionary Online*).

the state. State reporting is used by UN human rights treaty bodies,[20] quasi-judicial bodies comprised of independent experts, and the UN Human Rights Council, a state-centric, politicised peer-review mechanism.[21] These mechanisms are generally co-operative rather than adversarial; they do not engage solely with legal questions and analysis, rather they engage frequently with broad human rights topics of relevance to people's day-to-day lives.

State reporting as a mechanism was once described as 'the weakest in the range of implementation techniques' in international human rights law.[22] However, Brett later argued that providing information to UN treaty bodies in state reporting can raise issues of concern that may not meet the definition of 'violations', such as proposed laws where preventive action is required.[23] Further, the introduction of the Universal Periodic Review (UPR) – the cornerstone of the UN Human

20 See, for example, Anne F Bayefsky (ed), *The UN Human Rights Treaty System in the 21st Century* (Kluwer Law International 2000). State reporting obligations are outlined in each of the core international human rights law treaties, see section six below for further discussion. Ordered by date of adoption by the General Assembly, the nine core treaties are: International Convention on the Elimination of All Forms of Racial Discrimination, opened for signature 21 December 1965, 660 UNTS 195 (entered into force 4 January 1969) (ICERD); International Covenant on Civil and Political Rights, opened for signature 16 December 1966, 999 UNTS 171 (entered into force 23 March 1976) (ICCPR); International Covenant on Economic, Social and Cultural Rights, opened for signature16 December 1966, 993 UNTS 3 (entered into force 3 January 1976) (ICESCR); Convention on the Elimination of All Forms of Discrimination against Women, opened for signature 1 March 1980, 1249 UNTS 13 (entered into force 3 September 1981) (CEDAW); International Convention against Torture and Other Cruel, Inhuman and Degrading Treatment or Punishment, opened for signature 10 December 1984, 1465 UNTS 85 (entered into force 26 June 1987) (CAT); Convention on the Rights of the Child, opened for signature 20 November 1989, 1577 UNTS 3 (entered into force 2 September 1990) (CRoC); International Convention on the Protection of the Rights of All Migrant Workers and Members of their Families, opened for signature 18 December 1990, 2220 UNTS 3 (entered into force 1 July 2003) (CRMW); International Convention for the Protection of All Persons from Enforced Disappearance, opened for signature 20 December 2006, 2716 UNTS 3 (entered into force 23 December 2010) (CED); Convention on the Rights of Persons with Disabilities, opened for signature 13 December 2006, 2515 UNTS 3 (entered into force 3 May 2008) (CRPD).

21 Human Rights Council, *Institution-building of the United Nations Human Rights Council*, HRC Resolution 5/1, Human Rights Council 9th meeting, UN Doc A/HRC/RES/5/1 (18 June 2007).

22 Dana D Fischer, 'International reporting procedures' in Hurst Hannum (ed), *Guide to International Human Rights Practice* (University of Philadelphia Press 1984) 165.

23 Rachel Brett, 'State reporting: an NGO perspective' in Anne F Bayefsky (ed), *The UN Human Rights Treaty System in the 21st Century* (Kluwer Law International 2000) 57, 57.

Rights Council's Institution-building package,[24] has reasserted the importance of state reporting in international human rights law. As a mechanism, the UPR has been found to have significant potential in influencing state behaviour.[25] Treaty body state-reporting mechanisms have also been identified as having potential to address human rights issues related to conflict, including by forging synergies with transitional justice mechanisms and broader post-conflict recovery policies.[26] Further, several domestic factors may positively affect the effectiveness of treaty body state reporting, particularly the mobilisation of domestic actors.[27] As such, UN state-reporting mechanisms warrant examination as a key tool in the monitoring of international human rights law.

In addition to considering reports from states, UN bodies are reliant on NGOs in terms of their understanding of the issues on the ground, and in the development of appropriate recommendations.[28] It has been established that the role of NGOs in UN human rights bodies requires further examination.[29] Previous studies have examined the role in UN state-reporting mechanisms and heralded their importance, but often

24 Human Rights Council, *Institution-building* (n 21 above).

25 See, for example, Rochelle Terman and Erik Voeten, 'The relational politics of shame: evidence from the universal periodic review' (2018) 13(1) Review of International Organizations 1.

26 Evelyne Schmid, 'Socio-economic and cultural rights and wrongs after armed conflicts: using the state reporting procedure before the United Nations Committee on Economic, Social and Cultural Rights more effectively' (2013) 31(3) Netherlands Quarterly of Human Rights 241.

27 Jasper Krommendijk, 'The (in)effectiveness of UN Human Rights treaty body recommendations' (2015) 33(2) Netherlands Quarterly of Human Rights 194.

28 Fiona McGaughey, 'The role and influence of non-governmental organisations in the Universal Periodic Review – international context and Australian case study' (2017) 17(3) Human Rights Law Review 421; Fiona McGaughey, 'Advancing, retreating or stepping on each other's toes? The role of non-governmental organisations in United Nations Human Rights treaty body reporting and the Universal Periodic Review' (2017) 35 Australian Year Book of International Law 187.

29 See, for example, Linda Camp Keith, 'Human rights instruments' in Peter Cane and Herbert M Kritzer (eds), *The Oxford Handbook of Empirical Legal Research* (Oxford University Press 2010) 353, 364, 372; Fiona McGaughey, *Non-Governmental Organisations and the United Nations Human Rights System* (Routledge 2021).

without empirical evidence.[30] More recent scholarship has provided empirical evidence of the influence of NGOs on both UN treaty bodies and the UN Human Rights Council's UPR, but without a specific focus on transitional justice.[31] Therefore, more research on this essential role for NGOs is required with regard to transitional justice. It is argued here that both local NGOs and UN human rights institutions have the potential to play a critical role in supporting transitional justice. However, the effectiveness of both will be enhanced when there is active engagement between the local and the global – when NGOs play a *glocalising* role. It is acknowledged that NGOs undertake a range of international advocacy activities, in addition to engaging in UN state-reporting mechanisms, but these are outside the scope of our research.

A note of caution with regard to some of the scholarship in this area is that NGOs or civil society are often understood as unified, homogeneous entities.[32] It has been argued that conceptions of civil society have been limited to human rights NGOs, and that a rigorous conceptualisation of the role that civil society plays in transitional justice processes has been lacking.[33] Two main categories of NGOs (rights and reconciliation) and one sub-category (victim and survivor) are posited in the section below on 'Transitional justice from below in Northern Ireland'. However, these categories offer a lens through which to analyse NGOs' glocalising role in transitional justice and are not utilised to obscure the rich diversity of organisations differing in purposes, religiosity, locality, target groups, membership and so forth.

---

30 In relation to the ICESCR Committee, see, for example, Michael Freeman, *Human Rights: An Interdisciplinary Approach* (Polity Press 2011) 152. More generally, see, for example, Laurie S Wiseberg, 'The role of non-governmental organizations (NGOs) in the protection and enforcement of human rights' in Janusz Symonides (ed), *Human Rights: International Protection, Monitoring, Enforcement* (UNESCO Publishing 2003) 347, 350; David P Forsythe, *Human Rights in International Relations* (Cambridge University Press 2006) 203–204; Suzanne Egan, 'Strengthening the United Nations Human Rights treaty body system' (2013) 13(2) Human Rights Law Review 209, 227.

31 Lawrence C Moss, 'Opportunities for nongovernmental organization advocacy in the Universal Periodic Review process at the UN Human Rights Council' (2010) 2(1) Journal of Human Rights Practice 122; Edward McMahon et al, *The Universal Periodic Review, Do Civil Society Organization – Suggested Recommendations Matter?* (Friedrich-Ebert-Stiftung, Dialogue on Globalization 2013) . McGaughey (n 29 above).

32 Fiona McGaughey, 'From gatekeepers to GONGOs: a taxonomy of non-governmental organisations engaging with United Nations human rights mechanisms' (2018) 36(2) Netherlands Quarterly of Human Rights 111.

33 Paul Gready and Simon Robins, 'Rethinking civil society and transitional justice: lessons from social movements and "new" civil society' (2017) 21(7) International Journal of Human Rights 956.

Method

This socio-legal study situated within a human rights normative framework adopts a case study approach. With Northern Ireland as our case study, we focus on the space between top-down and bottom-up approaches, in which local NGOs have the potential to play an important intermediary role.[34] Case study research involves the detailed examination of a single example of a phenomenon.[35] While each case will have unique elements, this method allows for the development of hypotheses that can be further examined in other similar cases. Case studies can be a useful method for the examination of causal mechanisms and modelling and assessing complex causal relations.[36] In particular, process tracing can be used in case studies to map out complex interactions and track causality.[37] Here, we analyse whether recommendations in UN reports can be traced to NGO influence, or whether there are other causal links, although this is not process tracing as a distinct method *per se*, but it was informative in developing the current methodology; in particular, using documentary analysis to trace the influence of NGO reports through two UN human rights mechanisms.

A limitation of case studies is that it may be prohibitively time consuming to undertake case studies for a large number of situations or events,[38] and here we use only one case study – Northern Ireland. Northern Ireland has been selected as an appropriate case to explore the potential glocalising role of NGOs in an expanded transitional justice process for several reasons. It has a well-developed NGO sector with 140 charities focused on 'Conflict resolution or reconciliation or the promotion of religious or racial harmony or equality and diversity.'[39] It is an interesting jurisdiction for a case study because there has been an absence of a coherent top-down transitional justice process,[40] meaning local NGOs have had to develop innovative bottom-up processes to address past violence with little institutional support.

---

34 The UN now tends to adopt the broader term 'civil society', as such the term civil society is used in the article where relevant.

35 Bent Flyvbjerg, 'Five misunderstandings about case study research' (2006) 12 Qualitative Inquiry 219.

36 Alexander L George and Andrew Bennett, *Case Studies and Theory Development in the Social Sciences* (MIT Press 2005) 21–22.

37 Ibid.

38 Lisa Webley, 'Qualitative approaches to empirical legal research' in Cane and Kritzer (eds) (n 29 above) 926.

39 Northern Ireland Council for Voluntary Action, 'State of the Sector: Profile' (2022).

40 See, generally, Lauren Dempster, *Transitional Justice and the 'Disappeared' of Northern Ireland: Silence, Memory, and the Construction of the Past* (Routledge 2019).

Whether to select one or more case studies varies by academic discipline, for example sociologists favour multiple case studies and anthropologists favour a single case study.[41] As legal research has traditionally used doctrinal rather than empirical methods,[42] there is no accepted standard for the selection of, or number of, case studies in legal research. The lack of generalisable findings can also be a limitation; however, case-study research is designed to focus in detail on one selected situation rather than to provide findings that are generalisable to other situations.[43] Nevertheless, it may be possible to reach general conclusions, depending on the selected case.[44]

A systematic documentary analysis of UN human rights body reviews of the United Kingdom of Great Britain and Northern Ireland (the UK) was carried out to establish whether transitional justice issues are included in the UN deliberations and recommendations.[45] Submissions made by NGOs were also analysed. For UN treaty bodies, NGO reports are listed on the relevant website and were sourced from there. For the UPR, a 'stakeholder summary report' is prepared by the UN which lists NGOs which made submissions. These were used to assess which NGOs are engaged and whether transitional justice issues and recommendations are put forward by them. This allowed us to identify which types of NGOs – rights-focused or reconciliation-focused – are actively engaged with UN reporting frameworks. This article discusses the consequences of such engagement for transitional justice and makes recommendations for how the 'glocalising' potential of NGOs could be better supported by international mechanisms in societies emerging from mass violence.

In this article, section two charts current understandings of NGOs in transitional justice and human rights. Section three discusses the glocalising role of NGOs in the space between the global and local, which is followed by context-setting in relation to the conflict and post-conflict situation in Northern Ireland in section four. An examination of transitional justice 'from below' is presented in section five, including the work of the two main categories for the purposes of this article – rights-focused NGOs and reconciliation-focused NGOs. Grassroots activities are juxtaposed with UN human rights bodies'

41 Ibid.

42 Terry Hutchinson and Nigel Duncan, 'Defining and describing what we do: doctrinal legal research' (2012) 17(1) Deakin Law Review 83, 85.

43 Webley (n 38 above).

44 R K Yin, *Case Study Research Design and Methods* 2nd edn (Thousand Oaks 1994).

45 The documentary analysis was informed by Grant's scholarship: Aimee Grant, *Doing your Research Project with Documents: A Step-By-Step Guide to Take You from Start to Finish* (Policy Press 2022).

engagement with transitional justice in Northern Ireland in section six. In particular, that section considers whether UN bodies' engagement with transitional justice – or lack thereof – is related to the work of NGOs. Section seven reflects on the key findings, with conclusions and recommendations presented in the final section.

TRANSITIONAL JUSTICE FROM ABOVE OR BELOW: A FALSE DICHOTOMY

Transitional justice 'from above' typically refers to legalistic measures involving international or national institutions,[46] such as criminal prosecutions. The concern with prosecutions has arguably increased since the introduction of the Rome Statute in 1998,[47] and the growth of the principle of universal jurisdiction. Concomitant with the growth of truth commissions and criminal trials, however, has been an awareness of the need for civil society participation in these mechanisms.[48] These types of measures continue to be recommended to societies emerging from mass violence, despite mixed evidence of their effectiveness in supporting wider reconciliation. 'From above' can also refer to recommendations and views issued by UN bodies – the quasi-judicial mechanisms considered in this article.

The limitations of what has been called 'legalism' in relation to transitional justice have been well established. McEvoy, writing in 2007, declared: 'The field of transitional justice is increasingly characterized by the dominance of legalism, to the detriment of both scholarship and practice'.[49] Criticisms of the top-down approach to transitional justice point to failures of formal mechanisms to achieve their transitional

46 Kieran McEvoy and Lorna McGregor, 'Transitional Justice from Below: An Agenda for Research, Policy and Praxis' in McEvoy and McGregor (eds) (n 5 above) 1, 5.

47 Rome Statute of the International Criminal Court (opened for signature 17 July 1998) 2178 UNTS 3.

48 See, generally, Wendy Lambourne, 'Outreach, inreach and civil society participation in transitional justice' in Nicola Palmer, Phil Clark and Danielle Granville (eds), *Critical Perspectives in Transitional Justice* (Intersentia 2012) 235; Kjersti Lohne, 'Global civil society, the ICC, and legitimacy in international criminal justice' in N Hayashi and C M Bailliet (eds,) *The Legitimacy of International Criminal Tribunals* (Cambridge University Press 2017); Sarah Williams and Emma Palmer, 'Civil society and amicus curiae interventions in the International Criminal Court' (2016) 1 Acta Juridica. See with regard to Kenya, Christine Bjorkt and Juanita Goebertustt, 'Complementarity in action: the role of civil society and the ICC in rule of law strengthening in Kenya' (2011) 14(1) Yale Human Rights and Development Law Journal 205.

49 Kieran McEvoy, 'Beyond legalism: towards a thicker understanding of transitional justice' (2007) 34(4) Journal of Law and Society 411.

justice aims and to build sustainable peace.[50] A number of scholars have therefore argued that transitional justice aims can be best achieved through bottom-up initiatives that emerge from local organisations that are connected to grassroots communities and are sensitive to local complexities.[51] They argue that the sustainability of transitional justice initiatives is dependent on local ownership and participation, without which they lack legitimacy and will be ineffective.[52] It has also been asserted that top-down legalistic approaches are overly structured and lack transformative potential,[53] and that transitional justice should pursue socially transformative aims.[54]

A number of locally led and/or locally implemented transitional justice processes have been studied, such as Rwanda's Gacaca courts and Northern Ireland's truth-telling and storytelling projects.[55] A common argument made in their favour is that, as locally implemented processes, they enjoy a legitimacy and a level of engagement with ordinary citizens that externally mandated legal processes often lack. Local NGOs often lead these processes, although some rely on traditional leadership structures such as tribal elders – depending on context.[56] However, local initiatives also have their limitations, including a vulnerability to being distorted by asymmetric power relations.[57] Without oversight, transitional justice may be misused by local power holders, a risk identified by Chakravarty with the Gacaca courts in

50 Patricia Lundy and Mark McGovern, 'Whose justice? Rethinking transitional justice from the bottom up' (2008) 35(2) Journal of Law and Society 265; Paul Gready and Simon Robins, 'From transitional to transformative justice: a new agenda for practice' (2014) 8(3) International Journal of Transitional Justice 339.

51 Aiken (n 5 above); Lundy and McGovern (n 50 above).

52 Ibid.

53 Gready and Robins (n 50 above).

54 See, for example, Gready and Robins (n 50 above); Kris Brown and Fionnuala Ní Aoláin, 'Through the looking glass: transitional justice futures through the lens of nationalism, feminism and transformative change' (2015) 9 (1) International Journal of Transitional Justice 127.

55 See Jessica Senehi, 'The role of constructive, transcultural storytelling in ethnopolitical conflict transformation In Northern Ireland' in Judy Carter, George Irani and Vamik D Volkan (eds), *Regional and Ethnic Conflicts: Perspectives from the Front Lines* (Routledge 2015) 227; see also Claire Hackett and Bill Rolston, 'The burden of memory: victims, survivors, storytelling and resistance in the North of Ireland' (2009) 2(3) Memory Studies 355.

56 Volker Boege, Anne Brown, Kevin Clements and Anna Nolan, 'On hybrid political orders and emerging states: what is failing – states in the Global South or research and politics in the West?' (2009) 8 Berghof Handbook for Conflict Transformation Dialogue Series 15, 15–35.

57 Adam Kochanski, 'The "local turn" in transitional justice: curb the enthusiasm' (2020) 22(1) International Studies Review 26.

Rwanda.[58] Local initiatives may also struggle to achieve societal level change if they work with only limited numbers of individuals and are not scaled up to the national level.

In fact, the debate over the relative merits of top-down and bottom-up initiatives may be a false dichotomy. Top-down transitional justice processes can achieve a much wider impact when they are complemented by grassroots activities.[59] Meanwhile, local-level transitional justice processes may in fact rely on the protection and resourcing provided by governmental and international support.[60] For example, Wahyuningroem notes that Indonesia benefited from practices of transitional justice pioneered elsewhere and from the support of international organisations and donors – but that, alongside this, domestic human rights NGOs played important roles as norm entrepreneurs, helping to change how the public felt about human rights and pushing the state to change its behaviour.[61]

Hence, top-down and bottom-up approaches to transitional justice are not mutually exclusive. Instead, as explored in the next section, they play an intermediary role between local and international systems as a distinct contribution that locally grounded, internationally connected NGOs can make to transitional justice processes. This contribution, we argue, can be made by NGOs regardless of whether they view their primary aim as promoting human rights or supporting social reconciliation.

THE SPACE IN BETWEEN: A UNIQUE GLOCALISING ROLE FOR NGOs

NGOs have a role to play in transitional justice both locally and at the international level. In terms of engagement with UN human rights bodies, there are only limited opportunities provided for 'consultation' with NGOs in article 71 of the UN Charter, and as such, NGOs began as almost extraneous to the international human rights system.[62] In the last 71 years however, they have become increasingly important. In 1994, the UN Secretary General noted that 'NGO involvement has not

58 Anuradha Chakravarty, *Investing in Authoritarian Rule: Punishment and Patronage in Rwanda's Gacaca Courts for Genocide Crimes* (Cambridge University Press 2015).

59 Aiken (n 5 above).

60 Lina Strupinskienė, '"What is reconciliation and are we there yet?" Different types and levels of reconciliation: a case study of Bosnia and Herzegovina' (2017) 16(4) Journal of Human Rights 452.

61 Sri Lestari Wahyuningroem, *Transitional Justice from State to Civil Society: Democratization in Indonesia* (Routledge 2019) 121.

62 Charter of the United Nations (opened for signature 26 June 1945) 1 UNTS 16.

only justified the inclusion of Article 71 … but that it has far exceeded the original scope of these legal provisions.'[63] NGOs play a significant role in international human rights law, including their role in drafting international laws, bringing strategic litigation cases, and providing alternative reports in state-reporting mechanisms.[64] The information they provide to UN bodies is well reflected in UN recommendations, such as the human rights treaty bodies' concluding observations and the Human Rights Council's recommendations in its UPRs.[65]

At a local level, NGOs also play an important role in holding governments to account, and the UN human rights system can provide support for their efforts. While much scholarship on NGOs has traditionally focused on large international organisations, and, indeed, many such NGOs play a critical role in transitional justice,[66] the role of domestic NGOs in international human rights law requires further analysis. Seminal works, such as those by Risse-Kappen, proposed a transnational model to explain NGO behaviour, focusing on international NGOs and networks, but with less recognition of domestic NGOs.[67] This emphasis on *international* networks and NGOs mirrors the initial focus at the UN, which was on engagement with international NGOs as reflected in article 71 of the UN Charter.[68] Significant change was effected in 1996 when ECOSOC Resolution 1996/31 ('Consultative relationship between the United Nations and non-governmental organizations') recognised in its preamble 'the need to take into account the full diversity of the non-governmental organizations at the national, regional and international levels'.[69]

More recently, scholars argue that *domestic* NGOs and other actors are significant and that international human rights law must be adapted locally. Simmons and Merry have proposed that, although a repressive regime may need to be addressed using transnational NGOs, in most

63 UN Secretary General, 'General Review of Arrangements for Consultations with Non-Governmental Organizations: Report of the Secretary-General'.

64 McGaughey, 'From gatekeepers' (n 32 above); McGaughey, 'The role and influence' (n 28 above).

65 Ibid both articles.

66 See, for example, the International Centre for Transitional Justice.

67 Thomas Risse-Kappen (ed), *Bringing Transnational Relations Back In: Non-State Actors, Domestic Structures and International Institutions* (Cambridge University Press 1995).

68 Charter of the UN (n 62 above) art 71.

69 Economic and Social Council, *Consultative Relationship between the United Nations and Non-Governmental Organizations*, ECOSOC Res 1996/31, 49th Plenary Meeting (25 July 1996).

states, domestic actors are the most significant.[70] Simmons argues that international human rights law is powerful in mobilising domestic NGOs in holding states to account and that this is an important counterbalance to the dominance of scholarship on mechanisms such as transnational alliances.[71]

This article relies on the work of Simmons and Merry, drawing on the finding that domestic NGOs often have local, indepth expertise on the situation on the ground and play a useful role as intermediaries adapting international law to the local context through 'localized globalism'.[72] This has led to the current authors' use of the term 'glocalisation', denoting a process whereby NGOs can translate international human rights law to the local situation in a top-down direction, whilst submitting information to the UN on compliance in a bottom-up direction. This two-way flow of information managed by NGOs contributes to a democratisation of the UN human rights system and to local and global governance.[73] Meanwhile it has been argued that the global governing environment is one where a range of actors work together, and where simplistic international/local dichotomies are not reflected in practice.[74] Hence we argue that NGOs should be acknowledged as often operating in the space between top-down and bottom-up approaches to transitional justice.

Democracy and accountable governance are seen as essential components of transitional justice.[75] There is empirical evidence of a robust correlation between strong and autonomous civil society and positive human rights indicators post-conflict.[76] Based on a 2007 study of 60 states that transitioned from authoritarianism or a communist past, Tusalem concluded that: 'The strength of civil society prior to transition and its density post-transition not only play a significant role in the deepening of political freedoms and civil

70 Sally Engle Merry, *Human Rights and Gender Violence: Translating International Law into Local Justice* (University of Chicago Press 2006); Beth Simmons, *Mobilizing for Human Rights: International Law in Domestic Politics* (Cambridge University Press 2009).

71 Simmons (n 70 above).

72 Merry (n 70 above).

73 See, for example, Peter Willetts, 'From "consultative arrangements" to "partnership": the changing status of NGOs in diplomacy at the UN' (2000) 6(2) Global Governance 191; Richard Falk, *On Human Governance: Toward a New Global Politics* (Pennsylvania State University Press 1995) 241–255.

74 Charlotte Ku, *International Law, International Relations and Global Governance* (Routledge 2013) 131.

75 Wendy Lambourne, 'Transitional justice and peacebuilding after mass violence' (2009) 3(1) International Journal of Transitional Justice 28.

76 Rollin F Tusalem, 'A boon or a bane? The role of civil society in third-and fourth-wave democracies' (2007) 28(3) International Political Science Review 361.

liberties among transitional citizens, but also lead to better institutional performance.'[77] Similarly, Jeffery et al have argued that 'transitional justice and civil society have always gone hand in hand'.[78] They argue for the need to acknowledge the critical roles played by religious civil society organisations in transitional justice, writing specifically with reference to the Solomon Islands, Timor-Leste and Bougainville.[79] Aiken has also outlined a model of transitional justice where local NGOs in post-conflict societies play an essential part in the 'social learning' required to move from hostility to reconciliation.[80] Hence, local civil society actors play a key role in implementing and expanding transitional justice processes at multiple levels of society.

The ability of NGOs to play a glocalising role in the wider field of peacebuilding has been recognised by scholars who point to the unique potential for local NGOs to create reconciliation processes that are tailored to a specific context.[81] Since the end of the Cold War, internationally funded peacebuilding programmes have increasingly engaged with local NGOs as they seek to strengthen civil society capacities to sustain the post-conflict peace.[82] This has been heavily influenced by Lederach's model of conflict transformation which relies heavily on the agency of local civil society leaders and is driven by critiques that externally designed peacebuilding measures are ineffective at best, and at worst, cause harm to the post-conflict society.[83]

However, it may be unrealistic to expect local NGOs to effectively take forward a society-wide programme of reconciliation and ensure a non-recurrence of violence without connecting to the high-level institutional support that can be provided by international reporting frameworks in the field of transitional justice. In societies emerging from mass violence, civil society may be deeply divided along ethnic, religious or ideological lines, with only a limited number of individuals

77 Ibid 361.

78 Renee Jeffery, Lia Kent and Joanne Wallis, 'Reconceiving the roles of religious civil society organizations in transitional justice: evidence from the Solomon Islands, Timor-Leste and Bougainville' (2017) 11(3) International Journal of Transitional Justice 378.

79 Ibid 378.

80 Aiken (n 5 above).

81 Lambourne (n 75 above); Brandon Hamber et al, 'Towards contextual psychosocial practice' in *Psychosocial Perspectives on Peacebuilding* (Springer 2015) 289–316; Ian Patrick, 'East Timor emerging from conflict: the role of local NGOs and international assistance' (2001) 25(1) Disasters 48.

82 Thania Paffenholz, 'Unpacking the local turn in peacebuilding: a critical assessment towards an agenda for future research' (2015) 36(5) Third World Quarterly 857.

83 Ibid.

interested in promoting reconciliation between groups.[84] In such cases, high-level international support may help to motivate greater efforts towards reconciliation among local civil society.

In a meta-analysis of civil society peacebuilding, Paffenholz found that, while reconciliation-focused NGOs tend to concentrate their efforts on 'socialisation' activities to change attitudes, this has little societal impact.[85] Instead, she found that 'advocacy' activities had much more impact, especially when local civil society engaged with institutional processes such as the drafting of peace agreements or constitutions. Similarly, Diltmann et al also conducted a meta-analysis that only found substantial support for the effectiveness of three local-level peacebuilding activities; in-group policing, peace messaging and advocacy.[86]

The research to date therefore suggests that the time may be ripe for locally based reconciliation-focused NGOs to move beyond their traditional focus on improving intercommunal relationships to adopt a more advocacy-focused role where they engage with political institutions and legal frameworks in order to pressure reluctant political elites to engage in a meaningful process of transitional justice that could support social reconciliation.

NORTHERN IRELAND – TROUBLES AND TRANSITIONAL JUSTICE

The origins of the conflict in and about Northern Ireland are complex and contested.

It has been argued that its deepest roots, like many ethnic conflicts around the world, lie in the experience of colonisation,[87] even though the conflict is rarely described as such and Ireland is generally absent from lists of colonised nation states.[88] The period of 'the Troubles' saw sustained violence by paramilitaries and state forces from 1969–1998.

---

84 Rachel Rafferty, 'Engaging with the violent past to motivate and direct conflict resolution practice in Northern Ireland' (2017) 35(2) Conflict Resolution Quarterly 197.

85 Thania Paffenholz (ed), *Civil Society and Peacebuilding: A Critical Assessment* (Lynne Rienner Publishers 2010).

86 Ruth Diltmann et al, 'Addressing violent intergroup conflict from the bottom up?' (2017) 11(1) Social Issues and Policy Review 38.

87 Amy Maguire, 'Self-determination, justice, and a "peace process": Irish nationalism, the contemporary colonial experience and the Good Friday Agreement' (2014) 13 Seattle Journal for Social Justice 537.

88 Marc Mulholland (ed), *The Longest War: Northern Ireland's Troubled History* (Oxford University Press 2002).

More than 3,700 people were killed and over 40,000 were injured.[89] Further legacies of the violence include mental health impacts, with Northern Ireland suffering significantly higher rates of suicide than the rest of the UK and Ireland.[90]

In the 1990s momentum gathered for a peace process that resulted in the 1998 Belfast Agreement (the Agreement).[91] All the main political parties in Northern Ireland, leaders of the paramilitary organisations and the British and Irish Governments were involved in negotiations. The resulting Agreement made arrangements for political power-sharing between the two ethno-national communities. It contained strong commitments to human rights, including incorporation into Northern Ireland law of the European Convention on Human Rights (ECHR), with direct access to the courts, and remedies for breach of the Convention. It also provided for the establishment of the Northern Ireland Human Rights Commission (NIHRC) and the Equality Commission (Northern Ireland), designed to ensure that there would be no return to the discrimination and human rights abuses of the past. The Republic of Ireland also committed to similar measures, including establishing a 'Human Rights Commission with a mandate and remit equivalent to that within Northern Ireland'.[92] The essential nature of the human rights obligations underpinning this peace agreement explains the concern of commentators with the implications of Brexit for Northern Ireland, particularly when it appeared that the UK Government might retreat from the ECHR.[93]

The Agreement text also makes multiple references to the desirability of reconciliation, and the responsibility of the signatories to work towards this goal, but specific provisions for addressing the painful legacies of the past were not agreed. No formal comprehensive transitional justice process has been implemented in Northern Ireland. Indeed, to date, no agreement has been reached at the political level

89 Michael McKeown, 'An examination of the patterns of politically associated violence in Northern Ireland during the years 1969–2001 as reflected in the fatality figures for those years' (CAIN 2009).

90 Michael W Tomlinson, 'War, peace and suicide: the case of Northern Ireland' (2012) 27(4) International Sociology 464.

91 Also known as the Good Friday Agreement: 'The Belfast Agreement' (United Kingdom Government 1998). See, generally, Ian Somerville and Shane Kirby, 'Public relations and the Northern Ireland peace process: dissemination, reconciliation and the "Good Friday Agreement" referendum campaign' (2012) 1(3) Public Relations Inquiry 231.

92 'Joint Statement of IHREC and NIHRC', Northern Ireland Human Rights Commission.

93 David Phinnemore and Katy Hayward, 'UK Withdrawal ("Brexit") and the Good Friday Agreement' (European Parliament 2017); Amy Maguire, 'Brexit creates a human rights crisis for Ireland' (*The Conversation* 28 March 2017).

as to the nature of the past conflict, and as a result there is fierce contestation over how past events should be remembered and over how various parties that perpetrated violence should be treated. Irish Nationalist parties have expressed broad support for a process of truth-telling, and they also often call for public inquiries into abuses of human rights by state forces.[94] Meanwhile British Unionist parties have tended to reject any process they view as an amnesty for paramilitaries and instead advocate strongly for a narrative where state forces were noble defenders of law and order while paramilitaries can only be understood as murderous criminals.[95] At the same time, while the British Government has voiced generalised support for truth recovery where it pertains to the actions of paramilitary organisations, it has also directly refused to sponsor inquiries into alleged state-sponsored abuse such as collusion of members of the police with Loyalist paramilitaries.[96] In this environment of contested and competing narratives, despite a legal definition of 'victim',[97] victimhood has no agreed popular definition in Northern Ireland, and victims of the violence are used, and also at times choose to mobilise themselves, in support of political causes.[98]

While formal inquiry mechanisms have on occasion been implemented to highlight human rights abuses by state forces,[99] it is less clear how justice can be achieved for victims of paramilitary violence as the perpetrators are often unknown and evidence can be insufficient for prosecution. Sinn Fein has called for a truth-telling process in place of prosecutions, but this has been largely rejected by victims from within the Protestant community who see it as a means to avoid justice for violent crimes.[100]

---

94 John D Brewer and Bernadette C Hayes, 'The influence of religion and ethnonationalism on public attitudes towards amnesty: Northern Ireland as a case study' (2016) 22(4) Nationalism and Ethnic Politics 393.

95 Cheryl Lawther, 'Denial, silence and the politics of the past: unpicking the opposition to truth recovery in Northern Ireland' (2013) 7(1) International Journal of Transitional Justice 157.

96 Ibid.

97 Victims and survivors are defined in s 3 of the Victims and Survivors (Northern Ireland) Order 2006.

98 Laura Graham, 'The "innocent" victims of the troubles and the enduring impediment to peace in Northern Ireland' (2014) 17(1) Shared Space: A Research Journal for Peace, Conflict and Community Relations in Northern Ireland 37; Marie Breen-Smyth, 'Hierarchies of pain and responsibility: victims and war by other means in Northern Ireland' (2009) 25 Trípodos, Facultat de Comunicació i Relacions Internacionals Blanquerna 27.

99 Dermot P J Walsh, 'The Widgery Inquiry' in *Bloody Sunday and the Rule of Law in Northern Ireland* (Palgrave Macmillan 2000).

100 Breen-Smyth (n 98 above).

There have been a number of attempts to agree formal transitional justice mechanisms for Northern Ireland, but a coherent approach has not yet been implemented. In 2014 the Stormont House Agreement was signed by the major political parties in Northern Ireland and the British and Irish Governments. This agreement proposed the creation of an Historical Investigations Unit to investigate unsolved killings and allegations of police misconduct during the Troubles; an Independent Commission on Information Retrieval to allow families to privately receive information about the deaths of their relatives; an Independent Oral History Archive to record personal experiences of the conflict; and an Implementation and Reconciliation Group to oversee these legacy processes and promote reconciliation. More recently the British Government has tabled a Legacy Bill that, if passed into law, will offer conditional amnesty in exchange for cooperation with truth recovery mechanisms – proposals that have been widely rejected across Northern Ireland society.[101] However, as of the date of writing, the Stormont House Agreement has not been implemented and the Legacy Bill has not passed into law, and instead Northern Ireland continues to have a disjointed and *ad hoc* approach to investigating deaths during the Troubles.[102]

Northern Ireland has been relatively successful at preventing a recurrence of conflict and inter-community conflict, but some level of violence has continued,[103] and the deeper goal of social reconciliation remains out of reach. A number of initiatives have attempted to gain consensus on how to deal with the legacies of the violent past, but to date none have fully succeeded. For example, from May to October 2018, the Northern Ireland Office ran a public consultation process to gather public input on how to address the legacies of the conflict in Northern Ireland. The call for submissions was accompanied by a consultation paper that outlined plans, previously agreed in the 2014 Stormont House Agreement,[104] to establish legacy institutions and support victims and survivors.[105] Mallinder's subsequent qualitative analysis of submissions made to the 2018 consultation by Unionist political parties and Unionist-aligned organisations revealed strategic resistance to approaches to the past that are shaped by

101 Ross McKee, 'NI Troubles: amendments on controversial legacy bill revealed' (*BBC News* 24 November 2022).

102 Northern Ireland Affairs Committee, 'Addressing the legacy of Northern Ireland's past: the Government's New Proposals (Interim Report)' (26 October 2020).

103 Dale Pankhurst, 'Northern Ireland terror threat downgraded but Brexit tensions and threats of renewed violence remain' (*The Conversation* 30 March 2022).

104 'A Fresh Start: The Stormont Agreement and Implementation Plan' (17 November 2015).

105 Northern Ireland Office, 'Consultation Paper: Addressing the Legacy of Northern Ireland's Past' (May 2018).

international human rights obligations, and the significance of the ongoing 'metaconflict' in Unionist narratives about the past.[106] The Northern Ireland Office later published an analysis of the consultation process,[107] but actual progress on dealing with the past continues to be stymied by the metaconflict. There is no better reflection of this reality than the frequency and duration of suspension of the devolved power-sharing institutions – the Northern Ireland Assembly is suspended at time of writing, and was suspended from 2017–2020. It is within the gaps left by an absence of institutional transitional justice that NGOs in Northern Ireland have sought to achieve a measure of social reconciliation at the level of local communities.

TRANSITIONAL JUSTICE FROM BELOW IN NORTHERN IRELAND

Since the 1990s, a plethora of NGOs have emerged in the space created by Northern Ireland's transition out of mass violence. Cochrane and Dunn's analysis during the 1990s of 'peace and conflict resolution organisations' summarised the approach of the Northern Ireland NGOs as one of addressing symptoms rather than causes.[108] It has been argued that these organisations in Northern Ireland played a role in the peace process by creating new social forces and promoting an inclusivist ethos that was adopted by elites.[109] They also provided the opportunity for people to move out of paramilitary activity and into community activism (and sometimes political parties).

Having previously worked in this field, one of the authors observes that many NGOs have been explicitly devoted to an aim of social reconciliation, by focusing on reducing hostility and building trust and empathy between members of the Nationalist and Unionist communities. This peacebuilding work is known locally as 'community relations', and the Northern Ireland Community and Voluntary Association currently lists 140 community relations organisations out of over 7,486 NGOs and voluntary groups in the region.[110] An example

106 Louise Mallinder, 'Metaconflict and international human rights law in dealing with Northern Ireland's past' (2019) 8(1) Cambridge International Law Journal 5, 19.

107 Northern Ireland Office, 'Addressing the legacy of Northern Ireland's past: analysis of the consultation responses' (July 2019).

108 Feargal Cochrane and Seamus Dunn, *People Power? The Role of the Voluntary and Community Sector in the Northern Ireland Conflict* (Cork University Press 2002).

109 Feargal Cochrane, 'Unsung heroes or muddle-headed peaceniks? A profile and assessment of NGO conflict resolution activity in the Northern Ireland "Peace Process"' (2001) 12 Irish Studies in International Affairs 97.

110 Northern Ireland Council for Voluntary Action (n 39).

of a community relations NGO contributing to social reconciliation is Towards Understanding and Healing (TUH). Its mission is to facilitate creative conversations and thinking that moves beyond personal and societal conflict.[111] This type of NGO can be categorised as a 'reconciliation NGO', following the distinction outlined by Beirne and Knox.[112] In the post-conflict period, these local NGOs have been resourced largely by external funding from the European Union and other major donors, such as the International Fund for Ireland and Atlantic Philanthropies.[113]

A much smaller number of organisations in Northern Ireland fall into the category of rights NGOs. These tend to take a more rights-based approach to transitional justice, advocating strongly for the importance of acknowledging state abuses of human rights and of achieving reparations for victims as understood by international law. Reconciliation NGOs often view these rights-focused NGOs as distinct and even oppositional to the relational aims of reconciliation, but this may be an unnecessary dichotomy.[114] Rights-based NGOs in Northern Ireland also rely on external funding to carry out their work and they are more likely to engage with international frameworks to leverage support for their aims. A prominent example is the Committee for the Administration of Justice (CAJ). The CAJ was established in 1981 and is an independent NGO affiliated to the International Federation for Human Rights. Its aim is to 'ensure the highest standards in the administration of justice in Northern Ireland by ensuring that the government complies with its responsibilities in international human rights law'.[115] The CAJ works with other domestic and international human rights groups and makes submissions to UN and European human rights bodies. We refer to this second category as 'rights NGOs'. They may rely on external funding to carry out their work and engage with international frameworks to leverage support for their aims.

A third sub-set of Northern Ireland NGOs whose mission overlaps with the broad definition of transitional justice are support groups for victims and survivors of the Troubles. We term these 'victims and survivors organisations' (VSOs). This includes both rights and reconciliation NGOs. This category has a somewhat unique status as the interests of victims and survivors, services for them and consultation

111 'Towards Understanding and Healing – Work Plan' (CAIN). There are many other important NGOs engaged in this work, but this one is offered as an example.

112 Beirne and Knox (n 15 above).

113 Sean Byrne, *Economic Assistance and Conflict Transformation: Peacebuilding In Northern Ireland* (Routledge 2018).

114 Beirne and Knox (n 15 above).

115 'Committee on the Administration of Justice – Promoting Justice/Protecting Rights', Committee on the Administration of Justice.

with them are provided for in the Victims and Survivors (Northern Ireland) Order 2006 as amended by the Commission for Victims and Survivors Act (Northern Ireland) 2008. This status is also reflected in the provision of funding specifically for victims and survivors by the Office of First Minister and Deputy First Minister in Northern Ireland. An amount of £50 million was allocated for victims and survivors for 2011–2015 through the Victims Support Programme (for groups) and the Individual Needs Programme (for individuals).[116]

Some of these victims and survivors organisations are focused on supporting mental health while others provide a platform for victim advocacy.[117] We categorise these as support VSOs and advocacy VSOs. A further distinction across VSOs is that some organisations are cross-community, providing services for both communities, while others are single identity, comprising members from a single community. Where victims' advocacy groups focus only on the interests of their own community, this can limit their capacity to contribute to social reconciliation which requires recognition of the harms suffered by all parties to the conflict.[118] An example of an NGO that supports victims on a cross-community basis, but as a result must avoid taking a stance on controversial issues relating to the past violence, is the WAVE Trauma Centre. WAVE was set up in 1991 and aims to 'offer care and support to anyone bereaved, injured or traumatised through the civil unrest in Northern Ireland, irrespective of religious, cultural or political belief. The philosophy and ethos of the organisation is one of inclusiveness.'[119]

The next section reviews the contributions made by Northern Ireland NGOs to international human rights bodies. We then analyse this in light of the above typography, identifying which types of NGOs do and do not contribute to these reporting mechanisms. From this, we discuss the potential to create an international transitional justice architecture that might encourage greater participation from a wide range of transitional justice-focused local NGOs and enable their glocalising potential, for the good of their society and for the sake of comparative learning across contexts affected by mass violence.

---

116 Michael Potter and Anne Campbell, 'Funding for victims and survivors groups in Northern Ireland NIAR 576-14' (Northern Ireland Assembly, Research and Information Service 2014).

117 Laura K Graham, *Beyond Social Capital: The Role of Leadership, Trust and Government Policy in Northern Ireland's Victim Support Groups* (Springer 2016).

118 Rachel Rafferty, 'Conflict narratives, action frames, and engagement in reconciliation efforts among community activists in Northern Ireland' (2020) 26(1) Peace and Conflict: Journal of Peace Psychology 9.

119 WAVE Trauma Centre, 'About Wave Trauma Centre'.

TRANSITIONAL JUSTICE FROM ABOVE IN NORTHERN IRELAND

Having discussed the 'from below' activities of NGOs in support of transitional justice in Northern Ireland, this section focuses on 'from above' activities by the UN human rights bodies in order to identify any intersection between the international architecture and local NGOs in this case. The UN human rights system was developed largely as a response to the atrocities of the Second World War, as reflected in the preamble to both the UN Charter[120] and the Universal Declaration of Human Rights.[121]

> We the peoples of the United Nations determined to save succeeding generations from the scourge of war, which twice in our lifetime has brought untold sorrow to mankind, and to reaffirm faith in fundamental human rights, in the dignity and worth of the human person ... (UN Charter preamble)[122]

As such, an interest in preventing conflict was central to the establishment of the UN and the UN human rights bodies. Post-conflict societies often lapse back into conflict;[123] therefore, we argue that a concern for transitional justice should be germane to the purposes of the UN and its primary purpose of maintaining international peace and security. Furthermore, UN human rights bodies specifically should have regard to transitional justice-related issues, as it is well established that conflict correlates robustly with the violation of human rights.[124]

Universal mechanisms for transitional justice have been the subject of criticism and debate. In terms of a 'top-down' approach, the flaws of existing universal mechanisms, such as the International Criminal Court, have been identified,[125] and it has been suggested by Ramji-Nogales that the UN could create a new body responsible for transitional justice.[126] Yet several bodies already exist. In addition to mainstream UN human rights bodies and the International Criminal Court, the UN Peacebuilding Commission has been in place since 2005. Part of its remit is to 'bring together all relevant actors to marshal resources and to advise on and propose integrated strategies for post-

120 Charter of the UN (n 62 above).

121 Universal Declaration of Human Rights (adopted 10 December 1948 UNGA Res 217 A(III) (UDHR)).

122 Charter of the UN (n 62).

123 See, for example, Stephen John Stedman et al, *Ending Civil Wars: The Implementation of Peace Agreements* (Lynne Rienner Publishers 2002).

124 UN General Assembly, 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression' A/HRC/35/22 (30 March 2017).

125 Ramji-Nogales (n 11) 3.

126 Ibid 70.

conflict peacebuilding and recovery'.[127] As such the UN Peacebuilding Commission could build on this remit to play a valuable role in linking international human rights architecture with national transitional justice processes and the grassroots transitional justice activities of local NGOs, in support of an overall goal of social reconciliation for post-conflict societies.

Resolutions A/RES/70/262 and S/RES/2282 state that the Commission would serve as a platform to convene all relevant actors within and outside the UN, including 'civil society, women's groups, youth organizations and, where relevant, the private sector and national human rights institutions'.[128] The purpose of bringing actors together includes the opportunity for them to provide recommendations and information to improve their coordination and to develop and share good practices in peacebuilding.

Further, a new UN Special Procedure for transitional justice was established – the Special Rapporteur on the promotion of truth, justice, reparation and guarantees of non-recurrence. Mr Pablo de Greiff from Colombia was appointed first Special Rapporteur for transitional justice and took up his functions on 1 May 2012; he was replaced by Mr Fabian Salvioli (Argentina) in May 2018.[129] Aside from the Special Procedure, the UN Peacebuilding Commission and the International Criminal Court, given the plethora of international, regional and domestic human rights mechanisms already in existence in most parts of the world,[130] we must question whether a new UN body as proposed by Ramji-Nogales is required. Whether such a body is required leads us to question what existing UN bodies are doing with regard to transitional justice.

The UN bodies that will be considered here are those charged with periodic review of states' human rights records. The first of these is the Human Rights Council, the UN's primary human rights body. It uses the UPR to review all UN member states' human rights records on all human rights issues every four-and-a-half years. It is a peer review mechanism, meaning that the review is carried out by states and is political in nature. The second set of UN bodies are the human rights

127 UN General Assembly, 'Resolution Adopted by the General Assembly on 20 December 2005' A/RES/60/180 (30 December 2005); United Nations Security Council, 'Resolution 1645 (2005)' S/RES/1645 (20 December 2005).

128 UN General Assembly, 'Resolution Adopted by the General Assembly on 27 April 2016' A/RES/70/262 (12 May 2016); United Nations Security Council, 'Resolution 2282 (2016)' S/RES/2282 (27 April 27 2016).

129 Special Rapporteur on the Promotion of Truth, Justice, Reparation and Guarantees of Non-Recurrence, United Nations Human Rights Office of the High Commissioner.

130 Acknowledging that two of the authors' current home of Australia has somewhat limited domestic protections and is not part of a regional human rights system.

treaty bodies. These are bodies established under the nine core human rights treaties, such as the International Covenant on Civil and Political Rights,[131] which monitor states' compliance with their obligations under the treaty. The review is carried out by the treaty body members – voluntary, independent experts not representing their states of origin. Treaty bodies are thus quasi-judicial rather than political in nature. The treaty bodies consider only the issues relating to their treaty, so the review is thematic (rather than universal like the UPR), for example focusing on children's rights,[132] women's rights,[133] economic social and cultural rights[134] and so forth.[135] The treaty bodies only review those states which have voluntarily ratified the treaty, although levels of ratification are currently high.[136] The timing of their reviews can be more sporadic, and states are less likely to comply with the treaty bodies' reporting processes than they are with the UPR.[137]

As well as receiving reports from states, both the UPR and treaty bodies receive information from NGOs, and there is evidence that the NGO reports can influence the recommendations made by UN bodies.[138] The UPR also draws on a report summarising UN information – for example from Special Rapporteurs and treaty bodies – on the state under review, and this has also been shown to influence UPR recommendations.[139]

UN human rights state reporting with regard to Northern Ireland takes place through the reviews of the UK. As only one of the regions in the UK, it is plausible that specific human rights issues in one of several regions could be overlooked in UN state-reporting processes. This is particularly true of the Human Rights Council's UPR which is a brief review based on three very concise reports for each state under review. However, Northern Ireland is also the region within the UK with a recent history of violent conflict and alleged widespread human rights abuses, both by non-state actors and the state. From a transitional justice perspective therefore, it arguably requires a disproportionate focus in UN human rights reviews. The following paragraphs consider the level of engagement with transitional justice issues in recent reviews

131 ICCPR (n 20 above).
132 CroC (n 20 above).
133 CEDAW (n 20 above).
134 ICESCR (n 20 above).
135 CAT; CRPD; ICERD; CRPD; CED; CRMW (n 20 above).
136 'Status of Ratification: Interactive Dashboard', Office of the High Commissioner for Human Rights.
137 McGaughey, 'Advancing, retreating' (n 28 above).
138 Ibid; McGaughey, 'The role and influence' (n 28 above).
139 Ibid.

by these two main UN human rights state-reporting mechanisms – the UPR and the UN human rights treaty bodies.

In the UK's UPR in May 2017,[140] out of a total of 227 human rights recommendations to the UK, states made only two recommendations clearly related to transitional justice in Northern Ireland.[141] These were as follows:

> 134.156 Increase the necessary resources to the service of the Coroner to allow him to carry out impartial, swift and effective investigations on all the deaths linked to the conflict in Northern Ireland (Switzerland); and
>
> 134.157 Continue negotiations on transitional justice issues and implement transitional justice elements of the Stormont House Agreement (Australia).

A third recommendation is also likely to be driven by concerns for transitional justice:

> 134.67 Provide reassurance that any proposed British Bill of Rights would complement rather than replace the incorporation of the European Convention on Human Rights in Northern Ireland law and acknowledging this is a primary matter for the Northern Ireland Executive and Assembly — that a Bill of Rights for Northern Ireland to reflect the particular circumstances of Northern Ireland should be pursued to provide continuity, clarity and consensus on the legal framework for human rights there (Ireland).

The topic of a Bill of Rights for Northern Ireland has been contested for many years and, despite strong support and recommendations from the NIHRC, no action has been taken. A letter from the NIHRC to the Secretary of State in 2008 noted that, based on extensive consultations, '[w]hile there is agreement on having a Bill of Rights for Northern Ireland, this process has shown that there remains a diversity of opinion on the contents of such a Bill'.[142]

In addition to the UPR recommendations above, there was one transitional justice-related comment (rather than recommendation) from Ireland. It noted that the UK had changed its position on some

140 UN Human Rights Council, 'Report of the Working Group on the Universal Periodic Review – United Kingdom of Great Britain and Northern Ireland' A/HRC/36/9 (11–29 September 2017).

141 Interestingly, there was a post-colonial recommendation from the Syrian Arab Republic which could arguably be relevant to transitional justice in Ireland and Northern Ireland (ibid): 'Apologize to the peoples and the countries it colonized or it attacked and provide financial compensation to the peoples of these countries.'

142 'Advice to the Secretary of State for Northern Ireland' (Northern Ireland Human Rights Commission 9 March 2021).

previous recommendations and welcomed its commitment to establish an institutional framework to address the legacy of the Troubles.[143]

States in the UPR have the opportunity to choose whether they accept recommendations in the weeks following the review. A 2014 study found that 74 per cent of all recommendations made in the UPR are accepted by the state under review.[144] While states do not currently 'reject' recommendations (although this was a practice initially), Human Rights Council Resolution 5/1 provides that states may 'note' recommendations. The UK accepted 96 recommendations and noted 131 recommendations, explaining that this meant that it has 'taken some steps but it is not fully implementing them'.[145] The three transitional justice-related recommendations listed above were noted, but not accepted by the UK.

Previous research has shown that key sources for the development of recommendations are the inputs provided to states in the form of the UN compilation report and the stakeholder summary report (including NGO submissions).[146] One case study suggests that the UN summary could be the more influential of the two reports.[147] Given the scant attention to transitional justice in the UK's UPR, could it be the case that the issue was also largely overlooked in the UN compilation and stakeholder summary reports? Did states fail to address the issue because stakeholders, and the UN, had failed to initially raise the issue?

Firstly, let us consider the UN compilation report. The coverage of transitional justice in this short report was not insignificant. Of the 80 paragraphs, six paragraphs were relevant to transitional justice and recommendations covered investigations into conflict-related human rights violations and addressing the legacy of violations and abuses committed during the Troubles. Other recommendations related to reparations, truth and justice initiatives, freedom of assembly (violations of the determinations of the Northern Ireland Parades Commission) and segregated education.[148]

---

143 UN Human Rights Council (n 140 above) para 29.

144 UPR-Info, 'Beyond Promises: the impact of the UPR on the ground'.

145 UN Human Rights Council, 'Report of the Working Group on the Universal Periodic Review – United Kingdom of Great Britain and Northern Ireland – Addendum: Views on Conclusions and/or Recommendations, Voluntary Commitments and Replies Presented by the State under Review' A/HRC/36/9/Add1 (7 September 2016).

146 McGaughey, 'The role and influence' (n 28 above); Moss (n 31 above); McMahon et al (n 31 McMahon).

147 McGaughey, 'The role and influence' (n 28 above).

148 UN Human Rights Council, 'Working Group on the Universal Periodic Review: Compilation on the United Kingdom of Great Britain and Northern Ireland. Report of the Office of the United Nations High Commissioner for Human Rights' A/HRC/WG.6/27/GBR/2 (22 February 2017).

Was the transitional justice content of this report reflected in the recommendations made by states? Largely it was not. One of the recommendations from the UN compilation report was drawn from the UN Human Rights Committee and used similar language to Switzerland in its recommendation (above), recommending 'independent, impartial, prompt and effective investigations are conducted on conflict-related serious human rights violations in Northern Ireland'.[149] Another recommendation from the UN compilation report referenced the Committee on Economic, Social and Cultural Rights which urged the UK to expedite the adoption of a Bill of Rights for Northern Ireland. However, the recommendation used different language from that in the recommendation from Ireland (see above).

Secondly, let us consider the stakeholder summary report, which is of particular interest given this article's focus on NGOs. Very few Northern Ireland-based NGOs made submissions to the stakeholder summary. In fact no local NGO from any category (rights, reconciliation or VSO) made a submission, and only a few international NGOs who have a branch in Northern Ireland or a remit involving raising issues relating to Northern Ireland, such as Edmund Rice International and Amnesty International, made submissions. The few Northern Ireland-based NGOs which were referred to in the summary report do not work specifically on transitional justice issues, from either a rights, reconciliation or victim support perspective, and were cited with reference to other issues, such as the Council for the Homeless. In the report, only two organisations are quoted with reference to transitional justice – Amnesty International and the NIHRC.[150] As such, the coverage of transitional justice in this short report was less extensive than that of the UN compilation report. This means that grassroots voices from Northern Ireland on issues related to transitional justice – and recommendations that they might have – are largely absent from the UPR.

Of the 116 paragraphs in this report, only two focused on transitional justice and both of these had a legalistic focus on human rights issues but did not contain reference to the relational considerations of social reconciliation. The NIHRC called for 'impartial, prompt and effective investigations [to] be conducted into all conflict related deaths in Northern Ireland with a view to identifying, prosecuting and punishing

149 Ibid para 30.

150 National Human Rights Institutions sometimes carry out consultations to inform such submissions but there is no reference to consultations in the NIHRC submission to the UPR.

perpetrators of human rights violations and abuses'.[151] Amnesty International 'expressed concern that there had not yet been any concrete movement to create a human rights compliant mechanism for investigating and remedying past human rights violations and abuses that occurred during decades of political violence in Northern Ireland'.[152] Again, there was one additional recommendation which related to a Bill of Rights for Northern Ireland, which could arguably be driven by transitional justice considerations.[153] The 'Discussion' section below includes reflections on this dearth of transitional justice issues in the stakeholder report.

The level of engagement with transitional justice in the Human Rights Council's politicised UPR is to be contrasted with that of the independent UN treaty bodies in their state-reporting mechanisms. In short, treaty bodies have engaged more extensively with transitional justice issues in reviews of the UK carried out during a similar time period to the UPR. The reviews by the UN Committee Against Torture and the UN Human Rights Committee took place during a similar time period.

The UK's compliance with the Convention Against Torture and Other Cruel, Inhuman or Degrading Treatment or Punishment has been more recently considered in June 2019. At the Committee Against Torture 66th Session, Ms Thynne for the UK highlighted that 'an end to non-jury trials in Northern Ireland [would be possible], when safe and compatible with the interests of justice'.[154] The Committee referred back to recommendation 23 from its previous concluding observations in 2013. That recommendation was entitled 'Transitional Justice in Northern Ireland' and stated as follows:

> The Committee recommends that the State party develop a comprehensive framework for transitional justice in Northern Ireland and ensure that prompt, thorough and independent investigations are conducted to establish the truth and identify, prosecute and punish perpetrators. In this context, the Committee is of the view that such a comprehensive approach, including the conduct of a public inquiry into the death of Patrick Finucane,[155] would send a strong signal of

151 UN Human Rights Council, 'Summary of other Stakeholders' Submissions United Kingdom of Great Britain and Northern Ireland: Report of the Office of the United Nations High Commissioner for Human Rights' A/HRC/WG.6/27/GBR/3 (27 February 2017).

152 Ibid para 69.

153 Ibid para 34.

154 UN Committee Against Torture, 'Sixth Periodic Report of the United Kingdom of Great Britain and Northern Ireland', Summary of 1743rd Meeting (8 May 2019), CAT/C/SR1743 (14 May 2019) para 11.

155 Patrick Finucane was a lawyer killed during the Troubles, see, for example: BBC News, 'Q&A: the murder of Pat Finucane'.

> its commitment to address past human rights violations impartially and transparently. The State party should also ensure that all victims of torture and ill-treatment are able to obtain adequate redress and reparation.[156]

In 2019, the Committee concluded that 'Transitional Justice in Northern Ireland' had 'not been implemented'.[157] The Committee noted with concern the lack of effective investigation of multiple allegations of killing, torture and ill-treatment in the context of the conflict in Northern Ireland, the lack of accountability for perpetrators or redress for victims.[158] The Committee also noted the UK's failure to establish an independent historical investigations unit capable of examining killings and cases of alleged torture or ill-treatment where a detainee was not killed.[159] The Committee made six recommendations to the UK in relation to these matters, with a notable focus on independent investigation of torture-related allegations, transparency in process, ensuring that state agents are not immune from accountability, respect for the roles of journalists and human rights defenders, and redress for victims.[160]

The UN Human Rights Committee reviewed the UK's compliance with the International Covenant on Civil and Political Rights[161] in 2015. Of the 25 substantive recommendations from the Committee, two related to transitional justice in Northern Ireland. These are detailed and robust recommendations, one entitled 'Accountability for conflict-related violations in Northern Ireland' and the other entitled 'Fair trial and administration of justice'.[162] The Committee's recommendations included investigations, prosecution and remedies for victims, including investigation of all outstanding cases. It also recommended adequate resourcing of the Legacy Investigation Branch and the Coroner's Court to review outstanding legacy cases effectively. This resonates with Switzerland's recommendation to the UK in the UPR. This is evidence of the complementarity of the UPR and treaty bodies, as anticipated in General Assembly Resolution 60/251 which

156 UN Committee Against Torture, 'Concluding Observations on the Fifth Periodic Report of the United Kingdom of Great Britain and Northern Ireland', adopted by the Committee at its 50th Session (6–31 May 2013), CAT/C/GBR/CO/5, 24 June 2013, para 23.

157 UN Committee Against Torture, 'Concluding Observations on the Sixth Periodic Report of the United Kingdom of Great Britain and Northern Ireland', adopted by the Committee at its 1754th Meeting (16 May 2019), CAT/C/GBR/CO/6, 7 June 2019, para 7.

158 Ibid para 40.

159 Ibid.

160 Ibid para 41.

161 ICCPR (n 20 above).

162 Only part of the second recommendation relates to Northern Ireland.

stated that the UPR should complement, rather than duplicate, the work of UN human rights treaty bodies.[163] This refers primarily to non-duplication of the state-reporting mechanism.[164]

Like the Committee Against Torture previously, the Human Rights Committee also recommended an official inquiry into the murder of Patrick Finucane.[165] Furthermore, the Human Rights Committee requested follow up information on two of its recommendations within one year, in accordance with rule 71, paragraph 5, of the Committee's rules of procedure. Of the two recommendations selected, one was on accountability for conflict-related violations in Northern Ireland, which is evidence of the seriousness of the issue for the Committee.

Therefore, it is clear that treaty body reviews of the UK engage more deeply with Northern Ireland transitional justice issues than does the UPR. Although there may be a number of reasons for this (as considered in the 'Discussion section below), it has previously been established that NGOs can influence recommendations in the UPR and in treaty bodies. There are indications that the opportunity for NGO influence is greater in treaty bodies.[166] Let us consider whether increased engagement by Northern Ireland NGOs could be part of the reason for this greater engagement with transitional justice.

In the UK's review by the Human Rights Committee in 2015, civil society organisations made submissions for the session.[167] These included domestic NGOs from Northern Ireland, most notably Relatives for Justice and the Committee on the Administration of Justice, both of which we categorise as 'rights' rather than 'reconciliation' NGOs due to their focus on achieving legal justice. Amnesty International's submission also engaged with transitional justice issues in Northern Ireland, again a rights-based NGO, although international rather than domestic. Some organisations made earlier submissions in order to influence the scope of the review – known as the development of the 'list of issues prior to reporting' (LOIPR). Several rights NGOs raised Northern Ireland transitional justice-related issues, including the

163 Resolution Adopted by the General Assembly: 60/251 Human Rights Council, GA Res 60/251, UN GAOR, 60th session, Agenda Items 46 and 120, UN Doc A/RES/60/251 (3 April 2006).

164 Felice D Gaer, 'A voice not an echo: Universal Periodic Review and the UN treaty body system' (2007) 7(1) Human Rights Law Review 121.

165 UN Human Rights Committee, 'Concluding Observations on the Seventh Periodic Report of the United Kingdom of Great Britain and Northern Ireland', CCPR/C/GBR/CO/7 (17 August 2015) para 8.

166 McGaughey, 'Advancing, retreating' (n 28 above).

167 See UN Treaty Body Database.

Committee on the Administration of Justice, Rights Watch (UK) and Amnesty International.[168]

The Committee Against Torture's review of the UK in 2019 also elicited engagement by several local rights NGOs from Northern Ireland on transitional justice issues.[169] For example, civil society organisations including the Committee on the Administration of Justice, Amnesty International, the Pat Finucane Centre and Relatives for Justice raised justice issues with the Committee Against Torture, including accountability for past human rights abuses, investigations into deaths and legacy issues emerging from the Northern Ireland conflict. Further, the NIHRC made submissions to the Committee regarding investigations into conflict-related deaths in Northern Ireland.[170] Therefore, we can conclude that domestic NGOs from Northern Ireland have engaged more with treaty bodies in recent reviews than they have with the Human Rights Council's UPR in the context of transitional justice. It can also be seen that rights NGOs, potentially overlapping with VSOs that have an advocacy focus (such as Relatives for Justice), have been active in contributing to international reporting mechanisms while reconciliation NGOs have not engaged with these processes.

DISCUSSION

The issues raised in the previous section require some further analysis, particularly the reasons for the lack of transitional justice issues reported to the UPR by stakeholders (including NGOs), the lack of transitional justice recommendations by states in the UPR, and the lack of engagement by reconciliation NGOs and VSOs in UN human rights mechanisms generally.

The reasons for the lack of transitional justice recommendations by states in the UPR could be manifold. In the case of Northern Ireland, it may be perceived that the conflict has largely been resolved and is not a contemporary human rights priority area. This is a perception with which we disagree. It is also the case that states approach the UPR with their own agenda items and, if transitional justice is not one of them, it may be overlooked. Further emphasis on transitional justice as a core agenda item requires consideration by states. UPR recommendations can be associated with follow-up bilateral support and technical assistance – the type of approach that could support transitional justice. Another consideration is that, as a peer-review mechanism, the UPR is political in nature. It has been established

168 Ibid.

169 See UN Treaty Body Database (n 167 above).

170 Ibid.

that states may avoid certain issues due to fear of reciprocity, and that states' recommendations can be politically motivated.[171] Conflict and transitional justice are inherently sensitive issues for many states and are infused with politically charged issues. However, in some states, such as Sri Lanka, transitional justice-related recommendations arising from the UPR are more common.[172]

It is arguable that the UK may be exacerbating this problem through a lack of focus on Northern Ireland in its UN human rights reporting. The Human Rights Consortium has criticised the UK's 'consistent failures' in reporting on human rights with regard to devolved regions.[173] This means that it is all the more important for local NGOs to provide information to UN human rights reporting mechanisms. From the recommending states' perspective, there may be some inherent reluctance to identify a Western democracy such as the UK as requiring the input of other states on transitional justice. Nagy recognises that 'transitional justice almost always applies to non-Western, developing countries'.[174] This phenomenon reinforces a characterisation of Northern Ireland as a largely unrecognised colonial setting.[175]

More broadly, we question whether states are making transitional justice-related recommendations in the UPR? A database of all UPR recommendations is populated and maintained by international NGO UPR-Info.[176] This allows for filtering of recommendations by category such as 'women's rights', 'poverty' or 'land', but unfortunately 'transitional justice' is not included as a category. A keyword search can be done using 'transitional justice', but results are limited to when states have actually used that exact phrase. A keyword search of the database shows that of a total of 90,938 recommendations as of July 2022, only 113 include the phrase 'transitional justice'. This potentially leads to a certain invisibility of transitional justice concerns within human rights reporting frameworks, which in turn undermines the ability to present

171 Edward R McMahon, 'Herding cats and sheep: assessing state and regional behavior in the Universal Periodic Review mechanism of the United Nations Human Rights Council' (University of Vermont 2010).

172 UN Human Rights Council, 'Working Group on the Universal Periodic Review: Draft Report of the Working Group on the Universal Periodic Review – Sri Lanka', A/HRC/WG.6/28/L14 (23 November 2017).

173 'Database of Recommendations' UPR INFO.

174 Rosemary Nagy, 'Transitional justice as global project: critical reflections' (2008) 29(2) Third World Quarterly 281, cited in Kristine Höglund and Camilla Orjuela, 'Friction and the pursuit of justice in post-war Sri Lanka' (2013) 1(3) Peacebuilding 300.

175 Amy Maguire, 'Contemporary anti-colonial self-determination claims and the decolonisation of international law' (2013) 22 Griffith Law Review 238.

176 'Database of Recommendations' (n 173 above).

holistic recommendations. This is despite the likelihood that many more recommendations could relate to transitional justice – they just have not been framed in that language.

That may provide some understanding of state behaviour, but the question remains as to why transitional justice-related issues were largely absent from the UPR stakeholder summary report to which NGOs made the most significant contribution. The list of submissions which formed the basis of that report reflect little representation from NGOs exclusively based in Northern Ireland, and there were even fewer submissions from organisations working on transitional justice issues in the region. The international NGO Amnesty International did include transitional justice in its submission, as did the NIHRC. Both had recommendations included in the stakeholder summary report. However, there were no submissions from reconciliation NGOs or VSOs.

The lack of engagement with UPRs by NGOs that have the potential to make an important contribution to transitional justice requires further research as it means the glocalising potential of these organisations is currently unfulfilled in the Northern Ireland context. Possible reasons include a lack of awareness of the UPR as a mechanism, lack of inclusion of these groups in government consultations and coalitions,[177] lack of resources for engagement, and restrictions on advocacy work as a result of funding terms and conditions.[178] The failure of *reconciliation* NGOs in Northern Ireland to engage with core UN human rights mechanisms may also be partly explained by the perception of incompatibility between the rights and reconciliation approaches to peacebuilding identified by Beirne and Knox. Their work points to a wariness about human rights among reconciliation NGOs, despite the potential the authors identify for synergy between the two fields.[179] More broadly, of course, it is important to acknowledge the highly politicised nature of rights debates in Northern Ireland.[180]

This potential synergy has also been identified by Parlevliet with reference to the nexus of human rights and peacebuilding:

> It has become increasingly clear that the relationship of human rights and peacebuilding is complex, dynamic, and context-specific. While this interface was long thought to be inherently conflictual, it has transpired

177 McGaughey, 'From gatekeepers' (n 32 above).

178 Beirne and Knox (n 15 above).

179 Ibid.

180 This is made clear by Mallinder in her analysis of Unionist responses to proposals for dealing with the past: Louise Mallinder, 'Metaconflict and international human rights law in dealing with Northern Ireland's Past' (2019) 8(1) Cambridge International Law Journal 5.

> that it contains potential for considerable synergy too. Peacebuilding and human rights can complicate and strengthen one another.[181]

It has been argued therefore that NGOs 'need to increasingly consider how "peace" as a concept might more effectively be incorporated into their human rights work'.[182]

However, currently, in the case of Northern Ireland, there is neglected potential for reconciliation VSOs to play a glocalising role in support of transitional justice as they are currently not engaging with international frameworks such as UPR. This non-engagement is most likely due to the rights-based and legalistic focus of these frameworks and the perception among local NGOs that this is not relevant to their focus on building relationships and supporting the mental health of victim-survivors. The absence of local NGOs in glocalising work is significant because of their direct contact with grassroots Northern Ireland,[183] meaning that international frameworks are not capturing the experience of transition on the ground and that communities may be unaware of the potential support available to them at the international level to raise issues of concern.

The engagement by treaty bodies on transitional justice issues in Northern Ireland is more promising. One reason for this increased focus may be that the treaty bodies are not political bodies – they are independent experts who do not represent their home countries. Another reason is that there is evidence of more local NGO involvement in the treaty body state-reporting mechanisms, although this relates particularly to rights-focused NGOs, such as the Committee on the Administration of Justice. Northern Ireland's reconciliation NGOs do not seem to be engaging with treaty bodies currently. Again, this is a missed opportunity for NGOs focused on reconciliation to play a glocalising role in support of transitional justice. International human rights treaties are seen as significant for protecting human rights and preventing conflict. The Special Rapporteur on transitional justice submitted a report to the UN General Assembly in October 2017 in which he recommended ratification and incorporation of international

181 Michelle Parlevliet, 'Human rights and peacebuilding: complementary and contradictory, complex and contingent' (2017) 9(3) Journal of Human Rights Practice 357.

182 Charity Butcher and Maia Carter Hallward, 'Bridging the gap between human rights and peace: an analysis of NGOs and the United Nations Human Rights Council' (2016) 18(1) International Studies Perspectives 81.

183 Rights NGOs may also have grassroots membership and contact but the nature of the community-based work of reconciliation NGOs makes this connection to grassroots stronger.

treaties and noted that human rights violations can fuel conflict.[184] This is a positive development, as the potential contributions of the UN treaty bodies and the UPR and other mainstream UN mechanisms have hitherto been under-considered in the UN approach to transitional justice.[185] Yet the fulfilment of this potential will also rely on successful engagement with local NGOs, and, as this case study of Northern Ireland indicates, work is needed to convince a much wider range of NGOs that engagement with such legalistic mechanisms can enhance their own potential to contribute to transitional justice and, ultimately, reconciliation.

CONCLUSION

While the distinction between the state-centric review at the UN level and grassroots activities at a local level is ultimately a false dichotomy, this divide seems to be operating in practice in the case of Northern Ireland, where local NGOs are largely absent from international reporting frameworks. In general, states are closely involved in grassroots activities through funding and other measures and civil society is closely involved in the UN in a form of dispersed global governance. NGOs have the potential to play a unique local-to-global or 'glocalising' role, but this only works if local NGOs are enabled and encouraged to engage at a global level. There is a lack of evidence of this taking place in the Northern Ireland case study, particularly in regards to the reconciliation and victim-support aspects of transitional justice.

While transitional justice mechanisms and processes should be 'precisely tailored to particular events and societies',[186] lessons can be learned from case studies, including the current study. As discussed above (under the heading 'The space in between: a unique glocalising role for NGOs'), previous research demonstrates that, while reconciliation-focused NGOs tend to concentrate their efforts on 'socialisation' activities to change attitudes, this has little societal

184 'Report of the Special Rapporteur on the Promotion of Truth, Justice, Reparation and Guarantees of Non-Recurrence' A/72/43099 (12 October 2017) 9.

185 As reflected in United Nations General Assembly, 'Promotion and Protection of All Human Rights, Civil, Political, Economic, Social and Cultural Rights, Including the Right to Development – Resolution adopted by the Human Rights Council' A/HRC/RES/12/11 (12 October 2009); UN Secretary General, 'Guidance Note of the Secretary-General – United Nations Approach to Transitional Justice'; UN Human Rights Council, 'Promotion and Protection of All Human Rights, Civil, Political, Economic, Social and Cultural Rights, including the Right to Development – Resolution adopted by the Human Rights Council' A/HRC/RES/21/15, 11 October 2012.

186 Ramji-Nogales (n 11) 3.

impact, but that advocacy activities have much more impact.[187] Other scholarship concludes that advocacy is one of the three most effective local-level peacebuilding activities.[188] Hence, the apparent lack of engagement in local-to-international advocacy by reconciliation and victims NGOs in Northern Ireland means NGOs are not maximising their effectiveness to contribute to transitional justice and, ultimately, reconciliation.

Concomitantly, at an international level it is clear that UN mechanisms have the potential to improve human rights situations on the ground. The UPR as the cornerstone of the Human Rights Council is a powerful tool, a mechanism with which all states engage and one which could do more to expressly consider transitional justice and make relevant recommendations based on both UN and local inputs. There is more evidence of NGO engagement with treaty bodies in the Northern Ireland case study and of more indepth engagement with transitional justice issues by treaty bodies.

Where possible, local NGOs must be involved in both grassroots activities and international monitoring via the UN in order to maximise their glocalising potential. Using international mechanisms to raise issues that become politicised and intractable at a local level can be an important part of the transitional justice approach, but, where only rights NGOs – and not reconciliation NGOs – engage, the results are likely to be skewed towards legalistic approaches to the detriment of a social reconciliation focus. Governments and funders should support the advocacy work of *all* NGOs to enable them to engage in UN mechanisms. Human rights should not just be the domain of lawyers and rights-based organisations.[189] For transitional justice to be fully effective, it is important that the voices of those working in and with communities are heard.[190]

187 Paffenholz (n 85).

188 Diltmann et al (n 86).

189 See, for example, Jim Ife, *Human Rights from Below: Achieving Rights through Community Development* (Cambridge University Press 2009).

190 John Paul Lederach, *The Moral Imagination: The Art and Soul of Building Peace* (Oxford University Press 2005) 56.

Northern Ireland
Legal Quarterly
Autumn Vol. 74.3 (2023) 510–527
Articles DOI: 10.53386/nilq.v74i3.1105

MacDermott Lecture 2023: Confounding the rule of law: conflating immigration, nationality and asylum in the UK*

Devyani Prabhat

University of Bristol

Correspondence email: devyani.prabhat@bristol.ac.uk

ABSTRACT

In this MacDermott Annual Lecture, Professor Devyani Prabhat looks at current developments in immigration, nationality and asylum law and evaluates these in terms of the core ingredients of the rule of law. Specifically, the focus is on two aspects of the rule of law as elaborated on by Lord Bingham in his classic exposition on the rule of law (2011.)[1] First, Bingham asserts that 'Questions of legal right and liability should ordinarily be resolved by application of the law and not the exercise of discretion' is a fundamental requirement of the rule of law; and, second, Bingham states that 'the rule of law requires compliance by the state with its obligations in international law as in national law'. The examples Professor Prabhat analyses in the lecture are that of the East African Asians who could not enter the United Kingdom (UK) with their British passports in the 1960s and 1970s; the *Windrush* generation and the hostile environment of immigration control; immigration control of European Union nationals in the UK after Brexit; developments in cancellation of British Citizenship; and new legislation and proposals on asylum in the UK. Do these changes in the scope and application of the law comply with the rule of law in general and with the two specific principles on 'law not discretion' and 'international law compliance' in particular? While answering this question, Professor Prabhat explains how the different legal categories in immigration, nationality and asylum are distinct but are often conflated or confused with each other. Executive discretion should be narrow for nationality and asylum matters to conform with international law, whereas it can be wider for immigration so long as principles of fairness and non-discrimination are adhered to in each instance.

Keywords: asylum; immigration; nationality; citizenship; rule of law; discretion; cancellation; Brexit; *Windrush*; Commonwealth.

* MacDermott Lecture on 'The rule of law and immigration, nationality, and asylum', delivered at Queen's University Belfast on 18 May 2023.

1 Tom Bingham, T*he Rule of Law*, Penguin Law 2011.

INTRODUCTION

It is a huge honour to be speaking here in Queen's University Belfast in the 25th anniversary year of the Good Friday Agreement. I want to thank you for inviting me to speak at such a special event. The legacy of Lord MacDermott after whom this annual lecture is named is of immense value. In 1957 Lord MacDermott gave the Hamlyn lecture on the topic 'Protection from power' on the importance of the rule of law and the protection of human rights from the abuse of power.[2] It therefore seems apt to talk about the rule of law in contemporary times as well. Today I talk about the rule of law specifically in the context of immigration, nationality and asylum. These migration-linked topics are key areas for determining membership of a society and are critical for understanding a society's commitment to fairness and non-discrimination. Despite the breadth of subject matter, I have decided to keep all of immigration, nationality and asylum in the mix for the talk today because all three are linked conceptually, and through past and contemporary developments and cases. The conflation of these areas gives rise to far-reaching consequences for people's rights and for those who defend these rights and has grave implications for the rule of law.

WHAT IS THE RULE OF LAW?

The rule of law is a much-contested term, but its meaning has been deftly explained by Lord Bingham in his famous lecture at Cambridge University in 2006. The Cambridge lecture was later published as an article and then as a popular book on the rule of law.[3] 'The core of the existing principle,' Bingham argued, was 'that all persons and authorities within the state, whether public or private, should be bound by and entitled to the benefit of laws publicly and prospectively promulgated and publicly administered in the courts.' Bingham identified eight central principles attached to the rule of law:[4]

1 The law must be accessible and so far as possible intelligible, clear and predictable.
2 Questions of legal right and liability should ordinarily be resolved by application of the law and not the exercise of discretion.*

2 See Review by C of the speech 'Protection from power under English law' by Lord MacDermott: 'Review' (1958) 21(5) (Sept) Modern Law Review 569–573.

3 Tom Bingham, 'The rule of law' (2007) 66(1) Cambridge Law Journal 67–85; and Bingham (n 2 above).

4 Asterisks added by author for ease of reference to the principles in particular focus in this piece.

3 The laws of the land should apply equally to all, save to the extent that objective differences justify differentiation.
4 Ministers and public officers at all levels must exercise the powers conferred on them in good faith, fairly, for the purpose for which the powers were conferred, without exceeding the limits of such powers and not unreasonably.
5 The law must afford adequate protection of fundamental human rights.
6 Means must be provided for resolving, without prohibitive cost or inordinate delay, *bona fide* civil disputes which the parties themselves are unable to resolve.
7 The adjudicative procedures provided by the state should be fair.
8 The rule of law requires compliance by the state with its obligations in international law as in national law.*

The rule of law also forms part of the Constitutional Reform Act 2005. Section 1 of the Constitutional Reform Act simply states that the 'existing constitutional principle' of the rule of law and the Lord Chancellor's 'existing constitutional role' are not 'adversely affected' by the Act. Through this negative wording, the section confirms that the rule of law is indeed a constitutional principle.

Although several of Bingham's eight points are pertinent to our discussion, I have selected two (marked with asterisks above by me) which are particularly applicable to examples of conflated migration-linked categories. These are the second and eighth principles about law not discretion for questions of legal right and liability and compliance with obligations in international law as in national law. I shall discuss these two at greater length. While immigration, asylum and nationality are linked matters, they do have differences in how they are legally conceptualised. Immigration is mostly about meeting conditions and the exercise of executive discretion on entry and stay conditions, while nationality and asylum-seeking are matters of assessing legal rights and obligations. Yet these are not so distinct in the United Kingdom (UK) in their operation. In many instances, the use of wide executive discretion, unchecked by judicial process, permeates nationality and asylum as well as immigration. In all three areas international law principles as part of domestic law have been affected or even deliberately excluded in recent political developments and legal changes.

CONFLATING ASYLUM AND IMMIGRATION: THE UKRAINIAN SPONSORSHIP SCHEME

Immigration, nationality and asylum are regulated largely by the same statutes. For example, the Nationality and Borders Act 2022 aims to 'Make provision about nationality, asylum and immigration', amongst other objectives. Asylum is a matter of right in international law. Article 14 of the 1948 Universal Declaration of Human Rights states: 'Everyone has the right to seek and enjoy in other countries asylum from persecution.' Yet it is increasingly resembling immigration in terms of being similarly subject to discretionary decision-making powers. For instance, Ukrainian families who come to the UK must apply for visas in advance, which is the immigration concept of seeking permission in advance of entry and not based on a universal international right to seek asylum. The Ukraine Sponsorship Scheme allows Ukrainian nationals and their family members to come to the UK, but they first need to arrange for a sponsor who can provide accommodation for a minimum of six months.[5] In this manner, what was essentially a matter of right has become conditional and dependent. Families and individuals are left at the mercy of sponsors. There are reports of predatory exploitation and people being rendered homeless at the end of the six-month period.[6] These developments demonstrate how a core principle of the rule of law about applying the law rather than the exercise of discretion (Bingham's principle 2) is undermined and international law is not being implemented in national law (Bingham's principle 8). The Ukrainian scheme is paradoxically presented as a shining example of how the UK can accommodate and welcome asylum-seekers promptly and without delay as they would not have to seek asylum on arrival.[7] Despite its flaws, arguably, Ukrainian asylum-seekers are better situated than other asylum-seekers who cannot avail of any special schemes. I shall return to the situation of other asylum-seekers later in this lecture.

CONFLATING NATIONALITY AND ASYLUM: THE EAST AFRICAN ASIANS IN THE 1960s AND 1970s

The confusion in legal status is not just limited to asylum-seekers and immigrants but also permeates nationality-holding and asylum-

5 Guidance: Apply for a visa under the Ukraine Sponsorship Scheme (Homes for Ukraine).

6 'Homes for Ukraine: housing scheme called danger to refugees' (*BBC News* 4 May 2022).

7 Kathryn Cassidy, 'Homes for Ukraine: one year on' (*UK in a Changing Europe* 14 March 2023).

seeking. In the past this happened to British citizens who held British passports in Uganda and Kenya in the 1960s and 1970s. East African Asians who were coming to the UK to flee persecution there from newly formed authoritarian governments were denied entry into the UK (and often detained) despite being British passport holders. How did this happen? To understand why they could not enter readily with their passports we need to delve into the days of the British Empire.

Free movement of people was there in theory in times of Empire for British subjects, but in practice it was not really an option for most people. Travel between colonies (mostly non-white and not independent regions) and dominions (settled places which became independent or autonomous earlier) was usually restricted in practice.[8] Global politics was also shifting. British dominion Canada passed its own Citizenship Act in 1946 and issued Canadian passports to include its own French-Canadian citizens.[9] Canada's initiative in controlling its immigration and naturalisation meant that each dominion could potentially break away from any common understanding of subjecthood for immigration purposes. Each could determine criteria for entry and residence on its own terms and regulate subjects from other parts of the Empire. This challenged British supremacy which was at the heart of Empire and was part of the British vision of the UK's place in the post-Wars world. To counter this threat, the British Nationality Act 1948 expressly welcomed all Commonwealth nationals and people from former colonies who wanted to work or settle in Britain. It created the new status of 'citizen of the United Kingdom and Colonies' (CUKC) for people born or naturalised in either the UK or one of its colonies. Provision was also made in certain circumstances for citizenship to be acquired by descent from a CUKC, or by registration. The law was declaratory in nature in the sense that it did not require any further action by the affected people to take effect.

At the time of independence of the East African countries in the 1960s, the nationality arrangements were such that most Asians became citizens of the newly independent country in which they were living, but there were still many who held British passports and continued to do so. After the independence of Kenya, the Africanisation programme of the Kenyan authorities resulted in the persecution of the minority Asian community in Kenya. Large numbers of Asians with British passports began to migrate to the UK. Very soon there was

8 Randall Hansen, *Citizenship and Immigration in Post-war Britain: The Institutional Origins of a Multicultural Nation* (Oxford University Press 2000). See also Devyani Prabhat, 'Unequal citizenship and subjecthood: a rose by any other name …?' 71(2) (Summer) Northern Ireland Law Quarterly 175–191.

9 Laurie Fransman, *Fransman's British Nationality Law* (Bloomsbury Professional 2011).

virulent and vitriolic political opposition to the arrivals of Black and Asian people. A Member of Parliament, Enoch Powell, made a speech (popularly referred to as the 'Rivers of Blood' speech) which was about immigrants overcrowding Britain. The speech illustrates the naked hatred of migrants which unfortunately still has its echoes in modern migration discourse in Britain. Powell said:

> We must be mad, literally mad, as a nation to be permitting the annual inflow of some 50,000 dependants, who are for the most part the material of the future growth of the immigrant descended population. It is like watching a nation busily engaged in heaping up its own funeral pyre.[10]

This hostility in open political discourse was also matched by the everyday experience of racism of Black and Asian people in finding housing or employment in the UK.

Unsurprisingly, the British Government tightened controls over immigration. The 1962 Commonwealth Immigration Act rendered Commonwealth citizens subject to immigration controls for the first time; it ended the right of automatic entry for Commonwealth citizens. Thus, the status of subjecthood was now detached from any substantive rights, such as a right to reside in the UK. Even if they were ordinarily resident, or had been, they were subject to a new system enabling deportation of those who had committed criminal offences. All these changes permitted enormous administrative discretion in determining who can enter and who can stay in the UK or who can continue to remain without being evicted. Crucially, the 1962 Act removed the right of entry of CUKCs whose passports had been issued by colonial authorities.

In 1968 the British Government then passed another immigration Act in just three days: the Commonwealth Immigrants' Act 1968.[11] The 1968 Act sought to prevent the re-entry of people from countries such as Uganda and Kenya. A citizen could henceforth only live and work in the UK if they, or at least one of their parents or grandparents, had been born, adopted, registered or naturalised in the UK. The change in law now excluded most non-white people who were more likely to be born overseas or linked to the UK through parents or grandparents born overseas. This rule excluded almost all the East African Asians who were at that time seeking entry to the UK. It did this without expressly adding race as a direct criterion for inclusion or exclusion, and this made it difficult to challenge on grounds of discrimination in law.

---

10 'Enoch Powell's "Rivers of Blood" speech', 20 April 1968.

11 Randall Hansen, 'The Kenyan Asians, British politics, and the Commonwealth Immigrants Act, 1968' (1999) 42(3) Historical Journal 809–834.

The situation in East Africa worsened in 1972 when General Idi Amin ordered all Asians out of Uganda. Some 28,600 out of the 50,000 British passport holders in Uganda came to Britain. They were denied entry or detained in the UK and neighbouring countries. In *East African Asians v United Kingdom* (1973),[12] the court gave a decision about East African Asians and held that East African Asians who were British passport holders should not be deprived of their right of entry on racial grounds. Rather than assessing nationality rights, however, the court held that the UK had acted incompatibly with article 3 (freedom from torture, inhuman and degrading treatment), article 5 (right to liberty), article 8 (the right to respect for private and family life) and article 14 (prohibition on discrimination) of the European Convention on Human Rights 1950 (ECHR).[13] The case is considered a landmark pronouncement on inhumane and degrading treatment. However, it is also, arguably, a failure to pinpoint how and why citizenship rights matter. It leaves us with an unanswered question: how did the right to enter the country of nationality, which is a fundamental property of citizenship, become detached, discretionary and ultimately deniable?

CONFLATING 'IRREGULAR' PERSONS AND NATIONALS: THE WINDRUSH GENERATION

There are other enduring confusions about nationality from times of decolonisation which have led to present-day problems. People arriving from the Caribbean Islands from the 1940s to the 1960s had no requirement in law to register or to seek any sort of settlement status. Years later, the hostile environment legislations of immigration control with their requirements of document checking have resulted in the harassment, as well as detention and deportation, of many who had arrived earlier. The Home Office has now apologised for many of these errors which resulted in loss of life and liberty.[14] Compensation has been awarded to some people, although there are complaints about the slow rate of compensation payment and the bureaucracy surrounding the process.[15]

---

12 [1973] 3 EHRR 76.

13 The Immigration Act 1971 and the BNA 1981: the cross-linking of right of abode with being free of immigration control means citizenship is now a racialised notion (linked with ancestry and bloodlines). From 1983 no more special protection for Commonwealth citizens (they must naturalise like anyone else) and no more birth citizenship (BNA 1981).

14 'Home Secretary apologises to members of Windrush generation' (*Gov.uk* 10 June 2019).

15 'Windrush Compensation Scheme: impact assessment' (Home Office 6 February 2020).

Like most of the other examples in this lecture, the *Windrush* generation suffered (and continues to suffer) from dehumanisation and expulsion from membership of British society.[16] The vast amount of executive discretion in determining membership of society has rendered law an oppressive presence for many, rather than serving as a protection of rights and a guarantor of fairness and non-discrimination.

CONFLATING SETTLED MEMBERS OF SOCIETY OR IRREGULAR PERSONS: EU NATIONALS AFTER BREXIT

Another legacy of the document-checking regime is the precarious legal status of European Union (EU) nationals after Brexit. The EU Settlement Scheme (EUSS) is the mechanism through which EU, European Economic Area (EEA) and Swiss citizens and their family members resident in the UK prior to 31 December 2020 have been able to apply to secure their status and rights in the UK.[17] Pre-settled status, also known as Limited Leave to Remain, is a temporary form of stay in the UK. It is valid for five years. The pre-settled status of millions of EU nationals and their family members opened in August 2018 but is due to expire in the second half of 2023. This creates uncertainty about their future in the UK. The UK's position was that pre-settled status cannot generally be extended and is not upgraded automatically. Therefore, a subsequent application for settled status (officially called 'indefinite leave to remain') had to be made before the expiry date of pre-settled status. However, this was only possible once the applicant had completed five years of continuous residence in the UK. Those who failed to make this application were at risk of losing their right to remain in the UK.

The Independent Monitoring Authority brought a judicial review proceeding against the Home Office in the High Court to challenge the requirement of a second application. The High Court ruled in December 2022[18] that applicants granted pre-settled status should not lose their rights of residence if they do not make another application for settled status. The court based its decision on article 13(4) of the Withdrawal Agreement 2020 which states that a right of residence can only be lost in very specific circumstances and could not just be lost through the expiry of a previously held status. The court held that settled status rights accrue automatically, without the need of a second application if

16 'Windrush lessons learned' Independent Review by Wendy Williams (Home Office 19 July 2018, last updated 31 March 2020).

17 See 'Apply to the EU Settlement Scheme (settled and pre-settled status)' for details of the Scheme.

18 *Independent Monitoring Authority v Secretary of State for the Home Department* [2022] EWHC 3274 (Admin).

the other conditions (such as the five-year residence requirement) are satisfied. While the case was a victory for EU nationals, it illustrates how EU nationals have experienced categorical exclusion in a manner like Commonwealth nationals. Like the earlier arrivals, EU nationals have also had to challenge the loss of their legal status in the courts.

These examples of exclusions and challenges in the past and in the present are not just only instances of unfairness, but also illustrate the underlying point that, while distinct in legal conceptualisation, immigration, nationality and asylum do not exist as separate spheres of operation of political power. The lines between those who belong and those who do not are being continually redrawn. Here in Belfast, we need to reflect on the special implications of these developments for Northern Ireland. For instance, in Northern Ireland, the Good Friday Agreement birthright provisions allow the people of Northern Ireland to identify and be accepted as Irish or British or both.[19] Northern Irish people can choose to be both Irish and British or exclusively Irish or exclusively British, but at birth they are attributed to be British until they decide to exercise a choice.[20] On reaching the age of majority, they can renounce their British citizenship through an administrative process and become Irish or they can continue to be British.[21] However, post-Brexit EU law rights now would attach from being Irish. EU law rights are often more generous for bringing in third-country national partners and deriving other benefits, which means those who would like to retain EU rights would have to undergo renunciation of their British citizenship for pragmatic reasons. The numbers of these renunciations in Northern Ireland have soared in recent times.[22] Apart from the administrative inconvenience to people, this is problematic because of the serious identity-linked nationality issues in Northern Ireland.

---

19 Good Friday (or Belfast) Agreement: 'it is the birth right of all the people of Northern Ireland to identify themselves and be accepted as Irish or British, or both'.

20 The presumption of Britishness at birth was challenged by Emma DeSouza who was born in Northern Ireland and was treated as presumptively British by the Home Office. She was denied an EEA residence card for her US-born husband on that basis. The dispute was settled in a different manner for the DeSouzas, but before the Upper Tribunal they lost the case. The Upper Tribunal found in favour of the Home Office's position that the Belfast Agreement did not supersede the BNA 1981 and, therefore, Emma DeSouza was British despite her genuine belief that she was only Irish. See Upper Tribunal (Immigration and Asylum Chamber) in *De Souza (Good Friday Agreement: nationality)* [2019] UKUT 355 (IAC).

21 Section 12 of the BNA 1981: 'A British citizen of full age and capacity may renounce their British citizenship and that, subject only to concerns about statelessness, the Secretary of State shall give effect to that renunciation.'

22 Luke Butterfly, 'UK immigration law prompted many to renounce British citizenship in Northern Ireland' *Irish Times* (Dublin 14 November 2022).

CONFLATING NATIONALS AND FOREIGNERS: CANCELLATION OF CITIZENSHIP

The way citizenship is attributed at birth in Northern Ireland also resonates with one of the core issues in citizenship-stripping. At the time that a person faces the cancellation of their British citizenship they must be holding an alternate nationality so that they are not rendered stateless (section 40, British Nationality Act (BNA) 1981). Quite often the issue is unclear whether there is another surviving citizenship, but, through their past connections with other countries or their heritage or ancestry, they may be attributed an alternate nationality. It is then possible to strip a person of British citizenship because they would not become stateless as a result. This scenario plays out in the *Shamima Begum* (2023) case.[23] Begum left the UK as a 15-year-old British schoolgirl for Syria in 2015. She was found in a camp in Syria some years ago. The Home Secretary soon removed her British citizenship, thereafter, arguing that she would not be left stateless as she was eligible for Bangladeshi citizenship through her ancestry. This eligibility is derived from a Bangladeshi statute for nationality which has nothing to do with British nationality laws or with Begum's own actions in acquiring any other nationality. This was despite Bangladesh asserting that Begum is not their citizen and would be put on trial with the death penalty as an option for terrorism should she ever enter Bangladesh.[24]

After protracted litigation surrounding several preliminary issues, Begum lost her appeal against cancellation at the Special Immigration Appeals Commission (SIAC).[25] Yet the court found 'credible suspicion' that Begum had been trafficked for the purpose of sexual exploitation, as her lawyers had argued. It also found that there were 'arguable breaches of duty' by state authorities in having allowed her to make the journey to Syria.

---

23 *Shamima Begum v Secretary of State for the Home Department*, SIAC, date of judgment 22 February 2023.

24 Mattha Busby, 'Shamima Begum would face death penalty in Bangladesh, says minister' *The Guardian* (London 4 May 2019).

25 In *Begum's Case* (n 23 above), the preliminary issues were about the alternate nationality and about whether she could be permitted to enter the UK to be present in her appeals proceedings. Begum was stripped of her citizenship while outside the country and thus first sought permission to enter to be present at her appeal. This permission was denied by the Home Secretary, and she litigated the impact of this decision on her ability to instruct her lawyers and participate in her trial, arguing that it violated her fair trail rights. She lost that round of litigation in 2022 when it went up all the way to the Supreme Court. Only after that did the case return to the SIAC for hearing on the substantive grounds.

In its decision, the SIAC states that the case is 'about fundamental principles, rights and obligations'. It also says that 'the rule of law is non-negotiable'. But, ultimately, it upholds the cancellation order.

The *Begum* case and other cancellation cases take place within an opaque setting where national security trumps most other concerns. Cancellation of citizenship is easily done in the UK by a simple executive order (Home Secretary's order). In the UK, ministers are given decision-making authority based on their collective responsibility to Parliament. They are presumed to have superior knowledge and expertise. The Home Secretary appears to have a nearly unlimited degree of discretion in cancellation cases, even when human rights are at stake. There is no judicial oversight at the point of cancellation. The Home Secretary assesses what is a threat to national security and then decides whether to cancel citizenship or not.

Affected individuals do not receive an opportunity to make representations prior to the decision being made. Indeed, at times, the person affected does not even have to be notified of the order. Moreover, challenging a deprivation order is difficult. Appeals are only possible after the order comes into effect at which point cancellation will have already taken effect. Most people are outside the country when their citizenship is cancelled and are therefore unable to attend any legal challenges to the cancellation. Even when a person does appeal, their appeal is heard in a special court (the aforementioned SIAC). The SIAC holds closed proceedings when required in national security interests and gives closed judgments where national security-related material is involved. Special advocates provide legal support to appellants, but they only share the gist of the case with their clients and cannot take instructions once they have had access to any sensitive material.

Apart from these hurdles for appellants, the judges ordinarily only apply a very light-touch standard of review in national security cases. However, section 6 of the Human Rights Act 1998 requires courts and tribunals to act compatibly with the rights found in the ECHR. As a result, when rights are at stake, courts usually apply the more searching proportionality analysis while reviewing ministerial discretion. They are supposed to examine both whether rights were considered and whether they were attributed appropriate weight by the decision-maker. An example of how this operates is seen in immigration law where a person may challenge their deportation from the UK based on their right to a private and family life (article 8 of the ECHR, for example: House of Lords in *Huang* (2007)[26] on article 8 and proportionality). Courts try to determine through proportionality analysis whether individual rights are sufficiently protected.

While proportionality analysis is about the balance of factors, in the

26 *Huang* (2007) UKHL 11.

context of Begum's appeal (and others like it), it ought to be possible for courts to engage in a deeper analysis of the Home Secretary's decision-making should the court have been so minded. Appellate courts (such as the SIAC) are not confined by the relatively narrow standards of review of decision-making that ordinarily apply in judicial review proceedings. Instead, they can undertake what is called a 'full merits review' of a case. While this does not empower them to simply substitute their own views for those of the decision-maker, they should examine how the rights components as well as other relevant information formed a part of the original decision. Considering this, the SIAC's reluctance to narrow down the Home Secretary's ministerial discretion in *Begum*'s case despite serious concerns about statelessness, issues of fair trial or (as in the latest round) trafficking issues, is surprising and raises concerns.

A wider implication of *Begum*'s case is that anyone with any other national connection is now at greater risk of losing their British citizenship and becoming effectively stateless. The situation singles out naturalised citizens and other second-generation migrants born as British in terms of their holding a less secure citizenship status in the country. The SIAC noted in its decision that:

> many right-thinking people in this country's Muslim communities (and beyond) feel that they are being treated as second-class citizens, and/or that their welcome is somehow contingent. The Commission has received a considerable body of evidence on that topic, and it raises important issues. It is not an answer to that concern to say that the Secretary of State has paid regard at a general level to inter-community relations or was given advice that the deprivation of Ms Begum was strongly supported by a majority of public opinion.[27]

It then says that it has seen closed evidence that such an issue has been duly considered by the Home Secretary (para 398). These contradictions cannot be explained to the public without transparency in proceedings.

Attribution of nationality, such as in the *Begum Case*, is simply a mechanism of avoiding creating statelessness in the eyes of the law, while people are still left stateless in reality as they have no effective nationality. In terms of citizenship-stripping, because people who might have an alternate nationality are more likely to be vulnerable to this measure – which is usually connected to national security but can be for a range of different conduct – also has specific relevance for places where people may have multiple nationality, such as in Northern Ireland. The situation has now taken a worse turn because an amendment to section 40 of the BNA 1981 has now made it possible to render naturalised citizens stateless while cancelling their citizenship.

27 *Begum* (n 23 above) para 397.

Now an order to deprive a person of their British citizenship can be made by the Home Secretary, if the Home Secretary is satisfied that:

- it would be conducive to the public good to deprive the person of their British citizenship status and to do so would not render them stateless; or
- the person obtained their citizenship status through naturalisation, and it would be conducive to the public good to deprive them of their status because they have engaged in conduct 'seriously prejudicial' to the UK's vital interests, and the Home Secretary has reasonable grounds to believe that they could acquire another nationality.

So international law to a large extent is not taken seriously in this area, and there is only lip service to the idea of statelessness. Instead, what we see is a view of citizenship as most protected for a mono-national British by birth whereas everyone else can be stripped for a wide variety of conduct. This creation of various tiers of citizenship is, in fact, the creation of second-class citizenship.[28] Ngai while studying the treatment of East Asians in the United States (US) writes that, when ethnic minority citizens continue to remain aliens over generations, they can be called alien citizens.[29] Ngai's work was in the context of migrants acquiring citizenship rights, but for cancellation the focus shifts to treating citizens as problems. The idea appears to be to simply prevent re-entry, and the restrictions act to banish people. Apart from the punitive effect of banishment, there is the symbolic effect of such laws which seem to make some citizenship-holding of lesser value than others and to foster an idea of lingering foreignness.

To summarise, the wide discretion of the Home Secretary for cancellation cases is compounded by the avoidance of international law duties such as preventing trafficking, avoiding statelessness and providing a fair trial. All of these undermine the rule of law, particularly Bingham's principles 2 and 8 of the rule of law.

CONFLATING ASYLUM-SEEKERS AND 'IRREGULAR' PERSONS: ILLEGAL MIGRATION ACT 2023

Now, as promised at the beginning of the lecture, let us return to the topic of asylum and the rule of law. We have already seen that there is a need for prior permission to enter, such as with a visa for Ukrainians, or to arrive through a safe and legal route as set out in

28 Linda Bosniak, 'Varieties of citizenship' (2006) 75(5) Fordham Law Review 2449–2453.

29 Mae M Ngai, 'American orientalism' (2000) 28(3) Reviews in American History 408–415.

the Illegal Migration Act 2023. These are not requirements which can be imposed on asylum-seekers in international law. A requirement of prior permission increases the risk of unsafe conditions for asylum-seekers. Wide discretion with executive authorities in decision-making violate international law principles about the right to seek asylum.

Prior to its becoming law, the UNHCR said that the Illegal Immigration Bill (as it then was) would clearly breach the Refugee Convention, stating: 'The effect of the bill (in this form) would be to deny protection to many asylum-seekers in need of safety and protection, and even deny them the opportunity to put forward their case.'[30] While Ukrainian asylum-seekers have specific schemes, other asylum-seekers are provided very few means (if any) of entry into the UK. There has been particular focus on the Home Secretary's admission in March 2023 that the Illegal Migration Bill (as it then was) might conflict with requirements of the ECHR. On the first page of the published law, the then Home Secretary, Suella Braverman, said she was 'unable to make a statement' that the Bill's provisions were 'compatible with the Convention rights'; a highly unusual situation for a country ostensibly governed under the rule of law

Meanwhile the Government is looking to outsource asylum control to other 'safe' third countries. Certain asylum-seekers would not be allowed to enter the UK at all if asylum-seeking is offshore. The plan is to send these asylum-seekers to countries like Rwanda instead. At present no such transfers have taken place because of legal challenges.[31] However, the observable trend is very clear: just as we see the emergence of tiers of citizens, we can also see tiers of asylum-seekers who are treated differently from each other.

CONFLATING THE PROCEDURES OF THE RULE OF LAW WITH OBSTRUCTIONS AND THE DEFENDERS OF THE RULE OF LAW WITH OBSTRUCTIONISTS

It appears there is contravention of principles of international law in core areas of law relating to migrants (including those who are citizens but have connections with other countries) and migration. A further hurdle in many instances is that people cannot challenge these laws or decisions through effective means as due process in courts

30 United Nations High Commissioner for Refugees, 'UK asylum and policy and the Illegal Migration Act'.

31 The Supreme Court of UK decided on 15 November 2023 that it is unlawful to send asylum seekers to Rwanda under the Government's Rwanda Plan as there is risk of persecution, torture, or death if they are sent back to their countries of origin from Rwanda (refoulement): *R (on the application of AAA and others) v Secretary of State for the Home Department* [2023] UKSC 42.

is compromised. As we have seen, one of the most important courts for immigration matters is the SIAC. This court was initially set up in 1997, mainly as a forum where foreigners could appeal against deportation orders.[32] Later, it took over other national security cases, such as detention of those who could not be deported and cancellation of citizenship. In its early years after formation, the legal community criticised the SIAC's use of secret evidence and proceedings which were closed to the public. Many lawyers who served as special advocates at the SIAC resigned in protest because they were unhappy with how it operated.

Since then, its procedures have been fine-tuned (for example, those who appeal to it must be given the gist of the evidence against them). Its remit has also been expanded to include more immigration and nationality issues. It is not unusual for the SIAC to disagree with the Home Secretary on human rights issues. For instance, in a 2010 case, the court decided that the UK could not deport several suspected terrorists to Pakistan, as they faced a real threat of torture there.[33] This was despite the UK receiving diplomatic assurances from Pakistan that they would not be subjected to torture.

In recent years, however, it appears that the tide has turned on the evaluation of human rights issues in the SIAC. Higher courts have been steering the SIAC away from reconsidering the Secretary of State's assessment of the country's national security needs using its own lens. One example is a 2021 case, *Secretary of State for the Home Department v P3*,[34] where a person was deprived of their citizenship and then denied entry to the UK by the Home Secretary. The SIAC overturned the Home Secretary's decision, so the Home Secretary appealed. The Court of Appeal sided with the Home Secretary, ruling that the specialist court could not substitute its own evaluation of the interests of national security. This is also close to the Supreme Court's view as well in an earlier round of *Begum*'s *Case* (2021).[35] Here, too, the Supreme Court said that on a deprivation appeal, the SIAC is not entitled to re-evaluate the Home Secretary's discretion by using its own standards of review. In the absence of proper judicial oversight, it is nearly impossible to correct for (or even know of) any mistake or potential misuse of executive power.

Undermining the lawyers who have taken on cases in immigration, nationality and asylum has become a common occurrence in recent times. Their lives have at times been in grave danger. A man visited the

32 Louise Loveluck, 'The Special Immigration Appeals Commission (SIAC) explained' (*Bureau of Investigative Journalism* 24 October 2012).

33 Decision dated 18 May 2010, Appeal no: SC/77/80/81/82/83/09.

34 [2021] EWCA Civ 1642.

35 [2021] UKSC 7.

offices of Duncan Lewis Solicitors in Harrow armed with a large knife and threatened to kill a member of staff because he blamed lawyers at the firm for preventing the removal of immigrants from the UK.[36] Just days earlier the then Home Secretary, Priti Patel, had claimed activist lawyers were frustrating the removal of refused asylum-seekers from the UK. Another former Home Secretary, Suella Braverman, also wrote to Tory Party members claiming 'an activist blob of left-wing lawyers, civil servants and the Labour party' had opposed legislative attempts to curb small-boat crossings in the Channel.[37] In a similar vein in March 2023 Prime Minister Rishi Sunak said 'lefty lawyers' were thwarting efforts to crack down on illegal migration. The Bar Council and the Law Society have made representations to the Prime Minister to stop making statements that place lawyers at risk.[38] Such sentiments could serve to endanger professionals who work in this sector, especially lawyers seeking to defend the rule of law.

CONFLATING ETHNICITY, RACIALISATION, ALLEGIANCE AND MEMBERSHIP: SOME FINAL WORDS

We have come to a new crossroads now, 25 years from the Good Friday Agreement. With Brexit, and with other developments, immigration and nationality and asylum laws act as litmus tests for how seriously the rule of law is taken in a society. These laws are about membership in a society; who is an insider, who is an outsider. These may signify hospitality to others or hostility toward them. Disregard for international law in these scenarios may be interpreted as a lack of due regard for others in the world. Many commentators have said that recent hostility toward the EU is a resurgence of nostalgia for the continuation of Empire, but hostile environment developments have also affected members of the former Empire.[39] Perhaps there

36 Diane Taylor, 'Man faces terror charge over alleged attack at immigration law firm' *The Guardian* (London 23 October 2020).

37 Jasmine Cameron-Chileshe, William Wallis and Delphine Strauss, 'Sunak attacks "lefty lawyers" amid criticism of small boats strategy' *Financial Times* (London 9 March 2023).

38 'Bar Council and Law Society warn that Prime Minister's attacks on immigration lawyers are misleading and dangerous' (*Electronic Immigration Network* 14 June 2022).

39 Brexit has been driven by nostalgia for a lost empire, lost prominence of Commonwealth and Britain's global position in it, according to many scholars. Nobel laureate Abdulrazak Gurna has said that the British empire is 'still important in Britain' and may well have played a part in the Brexit vote. Afua Hirsch writes in her book *Brit(ish): On Race, Identity and Belonging* (Jonathan Cape 2018) 270 that 'the ghosts of the British Empire are everywhere in modern Britain, and nowhere more so than in the dream of Brexit'.

are pragmatic advantages in the continuation of relationships of exploitation, but such exploitation should never be justifiable in the name of law.

Relationships of exclusion are not just legacies of the past; there are also significant differences from how the state power would expand previously in terms of colonisation and how it operates now. In the past, colonisation involved the expansion of the jurisdiction of the state through the bodies of subjects. In the famous *Calvin Case* (1608) the Scottish-born subject was within the jurisdiction of the English King and could access estates in England. As a Scot, he could not legally own English land. Yet, through the concept that allegiance was tied to the person of the king, rather than to the kingdom itself or its laws, *Calvin's case* established that a child born in Scotland, after the Union of the Crowns under King James VI and I in 1603, was considered under the common law to be an English subject and entitled to the benefits of English law.[40] But now with developments, such as in cancellation of citizenship, there is a rollback of jurisdiction from the bodies of citizens. Citizens can be simply disowned and exiled. Such exile is usually linked to the concept of allegiance. For instance, in *Pham v Secretary of State for the Home Department* (2018), Arden, LJ, said (para 51 of the judgment):

> In the present case, the appellant has over a significant period of time fundamentally and seriously broken the obligations which apply to him as a citizen and put at risk the lives of others whom the Crown is bound to protect. I do not consider that it would be sensibly argued that this is not a situation in which the state is justified in seeking to be relieved of any further obligation to protect the appellant.[41]

It is hard to assess the objectives or efficacy of counter-terrorism measures such as cancellation of citizenship as very few hard facts are known. Perhaps the modern state is looking for a quick-fix pragmatic remedy to counterterrorism. Perhaps jurisdiction change could be because of the growth of human rights in the interim years from colonisation and the two World Wars to the present day, and now there is a pushback by the nation states to emerge stronger. Overturning these human rights frameworks which act as constraints over state action can be a show of nation-state power.

These are obviously broad issues we cannot resolve here in a single lecture, but categorical exclusion of foreigners present in our midst is often an indicator of rising national fervour and a return to ethnicised and racialised notions of national belonging. With hostile environment legislation it has now become the duty of individuals who have access

40 (1608) 77 ER 377.

41 [2018] EWCA Civ 2064.

to resources such as housing (landlords), health care (medical services staff) and employment (employers) to monitor the immigration status of others. Indeed, finding who is the 'other' has become the duty for all of us. As poet Alberto Rios has written: 'The border used to be an actual place, but now, it is the act of a thousand imaginations.'[42] This is all the more reason that we need to reflect on the principles of the rule of law, and why these should provide a structural framework of how we evaluate legal developments today. The principles apply irrespective of what specific views we hold on the various issues of migration that I have spoken about today.

CONCLUSION

It has been a pleasure to speak to you and to return to Belfast and to QUB where I gathered empirical data for my PhD several years back. That research led to my first book, *Unleashing the Force of Law*,[43] which eventually won the Peter Birks prize from the Society of Legal Scholars, UK, and Ireland. I owe a debt to this community and have a personal connection, and by inviting me to your midst for this conversation today you made me welcome once again. The core message of this lecture is that the grave implications for the rule of law intensify when different legal categories in immigration, nationality and asylum are collapsed into one. These ought to be considered legally distinct, but politics often conflates these and renders every legal status precarious. Executive discretion should be narrowly tailored for nationality and asylum matters to conform with international law, whereas it can be wider in scope for immigration so long as principles of fairness and non-discrimination are adhered to in each instance. Only then can there be conformity with the basic principles of the rule of law.

42 The Border: A Double Sonnet, Alberto Rios.

43 Devyani Prabhat, *Unleashing the Force of Law: Legal Mobilization, National Security, and Basic Freedoms* (Palgrave 2016).

Northern Ireland
Legal Quarterly
Autumn Vol. 74.3 (2023) 528–558
Articles DOI: 10.53386/nilq.v74i3.1119

The importance of being relational: comparative reflections on relational contracts in Australia and the United Kingdom†

Jessica Viven-Wilksch*

University of Adelaide, Australia
Correspondence email: jessica.viven-wilksch@adelaide.edu.au

ABSTRACT

The notion of the relational contract, previously limited to academic circles, is now being articulated by some courts. The consequences are threefold. Firstly, these judicial decisions are challenging the conception of agreements in the common law. Secondly, these decisions acknowledge the particularity of some long-term commercial relationships and shift the spotlight onto the relations of the parties. Thirdly, they are being used to integrate obligations to act in good faith. This article will show how these decisions implement the developing theory of relational contracts. The article will discuss recent developments in the United Kingdom and Australia and reflect on the parallel course the two jurisdictions are taking. By providing a bird's-eye view of normative changes affecting some long-term transactions, the aim of the article is to reflect on how contract law is being reshaped by the recognition that, in some contracts, the relationship, not self-interest, is the vital thing, demonstrating a move away from traditional contract law theory.

Keywords: relational contract; good faith; legal theory; morals; implied terms.

INTRODUCTION

There is a wind of change blowing on the traditional conception of contracts in the common law. The notion of the relational contract, previously limited to academic circles, is slowly beginning to be articulated by some courts. This notion is summarised in a recent article by Zoe Gounari as:

† First published in NILQ 73.AD2 94–124 6 October 2022.

* Lecturer, Adelaide Law School, University of Adelaide. I am very grateful for the comments and suggestions received on earlier versions of this article from the reviewers, Professor Matthew Harding, Professor Jeannie Patterson, Professor Severine Saintier, Mr David Christie and the Brown Bag collective from Adelaide Law School. I also thank Kate Leeson for her careful proofreading. All errors are mine.

> at its core contract is a relationship, albeit one with legal force, and requires parties' adherence to values conducive to a relationship not only for it to come about in the first place, but also for it to be successful down the line.[1]

Until recently, relational contract theory was just a controversial academic topic. Yet, over the last decade, the notion of the relational contract has become more prominent in judicial decisions in the United Kingdom (UK). Earlier decisions, while alluding to relational contracts, had yet to fully impact contract law doctrine in the common law. Then, in March 2019, in *Bates v Post Office Ltd*,[2] the Queen's Bench Division of the UK High Court of Justice explicitly stated that relational contracts do exist and that they can be defined through a list of criteria. Justice Fraser also applied his reasoning to find the contract in dispute was relational and implied obligations of good faith which bound the Post Office. The judge articulated good faith and relational contracts as two faces of the same coin. While a first instance decision, *Bates* matters as it emphasises the relevance and importance of the relational character of the contractual relationship. The consequences are threefold. Firstly, it challenges the classification of agreements in the common law. Secondly, it supports the argument that in some contracts the long-term nature to some extent, but in particular the nature of the project and of the relationship, bear legal consequences, thereby shifting the spotlight onto the relations of the parties in that classification. Thirdly, it is being used to integrate obligations to act in good faith. While the parties to the dispute in *Bates* later settled, this case remains an important development. Not only does it further advance the discourse on the relational contract, but it is also representative of a new wave of decisions in the UK acknowledging that there is more to an agreement than its written terms, warranting a contextual approach to the interpretation of contracts, from their terms to their enforcement. While a seminal decision, the question remains: is this more than a blip in time?

The aim of this article is to argue that the decision in *Bates* is part of the recognition by the judiciary of the relevance of context in contract law theory and contract law practice. Relational contracts place the emphasis on the circumstances of the parties leading to and during the performance of the contract, while remaining true to the written terms of the agreement. Context can be hard to determine, and this is one of the reasons why the notion of relational contracting is in itself controversial. During and beyond the COVID-19 pandemic, the need to renegotiate contracts and to compromise to ensure agreements

1 Zoe Gounari, 'Developing a relational law of contracts: striking a balance between abstraction and contextualism' (2020) 41(2) Legal Studies 176, 182.

2 *Bates v Post Office Ltd* (No 3) [2019] EWHC 606 (QB).

can be completed is becoming paramount to help economies recover. This is why this article argues that it is important to learn from recent legal developments in this area to facilitate trade and collaborative contracting.

The importance of context in interpreting contracts became even more prominent in 2020, when the World Health Organization declared the world was facing a global pandemic.[3] Within weeks, the Australian Federal Parliament released good faith principles for commercial tenancies whereby the landlord and the tenant were to discuss ways to maintain the tenancy rather than terminating for lack of funds due to the effects and uncertainties created by the pandemic. In 2020, at the height of the first wave of COVID-19, the spirit of the relational contract was also present in the code of conduct of commercial leasing which contained good faith principles.[4] This code came into effect in all states and territories from 7 April 2020 (being the date that National Cabinet announced a set of principles to guide the code to govern commercial tenancies affected by the COVID-19 pandemic), following legislation in the states and territories to implement the code.[5] Interestingly, the code's wording reminds us of the relational nature of commercial leasing contracts:

> Landlords and tenants share a *common interest in working together, to ensure business continuity*, and to facilitate the resumption of normal trading activities at the end of the COVID-19 pandemic during a reasonable recovery period … Landlords and tenants will *negotiate in good faith* … Landlords and tenants will act in an *open, honest and transparent manner*, and will each *provide sufficient and accurate information* within the context of negotiations to achieve outcomes consistent with this Code … Tenants must remain committed to the terms of their lease, subject to any amendments to their rental agreement negotiated under this Code.[6]

There is no real data on how many commercial tenancies were saved by this set of measures. Such data may not be gathered as this was only

---

3 World Health Organization, 'WHO Director-General's opening remarks at the media briefing on COVID-19 – 11 March 2020'.

4 Australian Government, National Cabinet Mandatory Code of Conduct: SME Commercial Leasing Principles during Covid-19 (2020) 2.

5 Leases (Commercial and Retail) COVID-19 Emergency Response Commercial Leases Declaration 2020 (ACT); Tenancies Legislation Amendment Act 2020 (NT); Retail and Other Commercial Leases (COVID-19) Regulation 2020 (NSW); COVID-19 Omnibus (Emergency Measures) (Commercial Leases and Licences) Regulations 2020 (Vic); proposed Commercial Tenancies (COVID-19 Response) Bill 2020 (WA); COVID-19 Emergency Response Act 2020 (Qld); COVID-19 Disease Emergency (Commercial Leases) Act 2020 (Tas).

6 Australian Government (n 4 above) 2 (emphasis added).

one of the measures adopted to bring the economy out of recession.[7] Notions of collaboration, cooperation and good faith are, however, evident in these principles. These notions have also appeared in codes of conduct legislated by the Federal Australian Parliament to regulate particular industries and their trading practices. Since 2014, first through specific provisions in the Franchising Code of Conduct, and later within the newly drafted Food and Grocery Code of Conduct, the Horticulture Code of Conduct and the Dairy Code of Conduct, good faith has been expressly legislated upon.[8] This article will demonstrate that the Australian legislative approach echoes the relational contract judicial developments in the UK, even though this has not been expressly pointed out.

This article will first provide a theoretical overview on relational contracting to provide a backdrop to the judicial and statutory developments later analysed. It will then provide a chronological development of judicial decisions in the UK and in Australia. This section will use the legal issues raised in *Bates* as a case study. The article will first review the context of the agreement, the terms of the contract and the relationship between the parties in the dispute. It will lay out the factual as well as the legal landscape of the case. To consider relational contracts is to go beyond the black letter of the law, and the article will illustrate that the context and conduct of the parties take a prominent place next to the written terms. The article will then reflect on the relevance of these developments in Australia and compare them with recent changes in the regulation of some industries. The contract regulatory landscape in Australia is different to the common law approach of the UK. Therefore, while case law is analysed, it is also relevant to consider regulatory reforms which have targeted agreements in particular industries. By providing a bird's-eye view of normative changes affecting some long-term contracts in the UK and in Australia, the aim of the article is to analyse the possible social and legal implications of the growing recognition of relational contracts.

---

7 Coronavirus Economic Response Package (Payments and Benefits) Rules 2020 (Cth); Coronavirus Economic Response Package (Payments and Benefits) Act 2020 (Cth).

8 Competition and Consumer (Industry Codes – Franchising) Regulation 2014 (Cth); Competition and Consumer (Industry Codes – Horticulture) Regulations 2017 (Cth); Competition and Consumer (Industry Codes – Dairy) Regulations 2019 (Cth); Competition and Consumer (Industry Codes – Food and Grocery) Regulation 2015 (Cth).

RELATIONAL CONTRACTS: A SOCIAL CONCEPT WITH LEGAL CONSEQUENCES

Reclassifying contracts using a spectrum

Contracts are based on the parties' intention to enter into an agreement, and their relations are dictated on their own terms. Freedom to contract and party autonomy have dominated contract law theory. Parties express their intention by exchanging promises and by regulating their own dealings in their own terms. This ideal scenario is illustrated by the idea that, according to economic theory, each party to a contract is a rational actor. A person decides to enter into a legal agreement if the advantages outweigh the costs. Therefore, the goal of the transaction is to see it successfully performed. In this theoretical situation, both parties gain from the transaction because they are 'self-interested egoists who maximise utility'.[9]

Contract law involves more than just protecting party autonomy.[10] The ideal scenario is not often realised in practice. This might be because there is an imbalance of power between the parties, or one party might be more knowledgeable than the other. Where a party acts opportunistically and seeks to take advantage of the other party, some limits are placed upon autonomy. Worthington identifies different types of constraints.[11] For instance, general constraints prevent a party from hiring an assassin. Perhaps most significantly, there are constraints in contract law itself, as illustrated by the importance of consent and legal interventionism to protect the idea that consent should be freely given. The doctrine of unconscionability is a particularly relevant example of this restraint.[12] This intervention is not only apparent in the development of vitiating factors including unconscionability but also in the limits placed on the exercise of contractual powers during performance and termination of the contract. There are several examples

9 Cento G Veljanovski, 'Economic approach to law: a critical introduction' (1980) 7 British Journal of Law and Society 158, 162.

10 This is demonstrated by the legal literature: see Reinhard Zimmermann, *Roman Law, Contemporary Law, European Law: The Civilian Tradition Today* (Oxford University Press 2001) 174.

11 Sarah Worthington, 'Common law values: the role of party autonomy in private law' in A Robertson and M Tilbury (eds), *The Common Law of Obligations: Divergence and Unity* (Hart 2016) 303–306.

12 Competition and Consumer Act 2010 (Cth), sch 2 Australian Consumer Law, ss 20–22; *Commercial Bank of Australia Ltd v Amadio* [1983] HCA 14; *Alec Lobb (Garages) Ltd v Total Oil* (Great Britain) Ltd [1983] 1 WLR 87; see also Ying Khai Liew and Debbie Yu, 'The unconscionable bargains doctrine in England and Australia: cousins or siblings?' (2021) 45(1) Melbourne University Law Review (advance copy).

of this phenomenon, including the implications of terms in fact[13] and in law,[14] the limits placed on the enforceability of exclusion clauses,[15] and penalty clauses[16] and the rejection in some countries, such as Australia, of the efficient breach.[17] Party autonomy is also limited by complaints of unfair dealing.[18] This shows that there are limits to how far a contractual party can exercise its autonomy. The foundations for such rejection can be found in the primacy of the promise, as well as the infiltration of moral values into the regulation of contractual dealings. How much these values impact on contract drafting, conduct of the parties and enforcement of agreements depends on the type of contract.

The term relational contract was coined by Ian Macneil.[19] His theory is based on two main pillars: the length of the contract and the relationship between the parties. According to him, the life of the contract is not entirely predictable; therefore, the agreement will always be, to some extent, incomplete. He adds that '[t]he more relational an exchange, the less likely the parties plan and allocate risks effectively'.[20] Macneil identifies 10 norms that define a contract as relational: (1) role integrity (requiring consistency, involving internal conflict, and being inherently complex); (2) reciprocity (the principle of getting something back for something given); (3) implementation of planning; (4) effectuation of consent; (5) flexibility; (6) contractual solidarity; (7) the restitution, reliance and expectation interests (the

13 *BP Refinery (Westernport) Proprietary Limited v Shire of Hastings (Victoria)* [1977] UKPC 13; *Marks and Spencer plc v BNP Paribas Securities Services Trust Company (Jersey) Ltd* [2015] UKSC 72; *Codelfa Construction Pty Ltd v State Rail Authority of NSW* (1982) 149 CLR 337.

14 *Liverpool City Council v Irwin* [1976] UKHL 1; *Commonwealth Bank of Australia v Barker* (2014) 253 CLR 169.

15 *Davis v Pearce Parking Station Pty Ltd* [1954] HCA 44; *Sydney City Council v West* (1965) 114 CLR 481; *Renard Constructions (ME) Pty Ltd v Minister for Public Work* (1992) 26 NSWLR 234; *Persimmon Homes Ltd v Ove Arup & Partners Ltd* [2017] EWCA Civ 373.

16 *Paciocco v ANZ Group Ltd* [2016] HCA 28 - 258 CLR 525; *Cavendish Square Holding BV v Makdessi* [2015] UKSC 67; see also Mary Arden and James Edelman, 'Mutual borrowing and judicial dialogue between the apex courts of Australia and the United Kingdom' (2022) 138 Law Quarterly Review 217.

17 *Tabcorp Holdings Ltd v Bowen Investments Pty Ltd* [2009] HCA 8 [13]; for a review of the UK, see also Solène Rowan, 'Abuse of rights in English contract law: hidden in plain sight? (2021) 84(5) Modern Law Review 1066, 1067.

18 Paul Finn, 'Fiduciary and good faith obligations under long term contracts' in Kanaga Dharmananda and Leon Firios (eds), *Long Term Contracts* (Federation Press 2013) 137.

19 Ian R Macneil, *The New Social Contract: An Inquiry into Modern Contractual Relations* (Yale University Press 1980) 10.

20 Robert A Hillman, *The Richness of Contract Law: An Analysis of Critique of Contemporary Theories of Contract Law* (Kluwer 1998) 256.

'linking norms'); (8) creation and restraint of power (the 'power norm'); (9) propriety of means; and (10) harmonisation with the social matrix, that is, with 'supracontract norms'.[21] These elements are at odds with the classical view of contracting, or 'egoist' contracting.[22] These norms echo notions of cooperation, communication and transparency between the parties. These parties are respectful, loyal and take into consideration the interests of the other party. To be clear, this does not mean that parties determine their actions based on the interests of the other party, ie a fiduciary relationship. Macneil has summarised his arguments in four main strands:

> First, every transaction is embedded in complex relations.
>
> Second, understanding any transaction requires understanding all essential elements of its enveloping relations.
>
> Third, effective analysis of any transaction requires recognition and consideration of all essential elements of its enveloping relations that might affect the transaction significantly.
>
> Fourth, combined contextual analysis of relations and transactions is more efficient and produces a more complete and sure final analytical product than does commencing with non-contextual analysis of transactions.[23]

Macneil's theory, although it was never intended to become one, has been fiercely criticised.[24] The categories laid out by Macneil are not finite. This means that not all commercial contracts will be relational.[25] The long-term nature of the contract is one but not the only criterion.[26] This relational approach has been criticised by those who consider that these standards decrease predictability, with the ultimate consequence of increasing transaction costs.[27] For some, it is one reason for the lack of doctrinal impact, in so far as that Macneil's notion of relational contract was developed not as a theory, but as a sociological contractual phenomenon. It has been described as a simple matter of 'spotlight

21 Ian Macneil, 'Relational contract theory, challenges and queries' (2000) 94 Northwestern University Law Rev 877, 879–80.

22 See Veljanovski (n 9 above).

23 Ibid 881.

24 For a review of the use of the theory, see Josetta McLaughlin, Jacqueline McLaughlin and Raed Elaydi, 'Ian Macneil and relational contract theory: evidence of impact' (2014) 20(1) Journal of Management History 44.

25 *Bates v Post Office Ltd (No 3)* [2019] EWHC 606 (QB) [705], [714].

26 *D&G Cars Ltd v Essex Police Authority* [2015] EWHC 226 (QB).

27 For a review, see James Gordley, 'The moral foundations of private law' (2002) 47 American Journal of Jurisprudence 1.

orientation',[28] emphasising the relationship rather than the exchange itself.

Tan, in his seminal article on relational contracts, provides a framework to grasp relational contract theory's possible doctrinal ramifications through three pathways. The first is re-interpretive relationalism, meaning other established concepts already reflects.[29] The second is re-orientative relationalism explained by Tan as involving 'intra-doctrinal salience and additional alteration'[30] using Leggatt J's judgment in *Yam Seng* as a prime example of this development.[31] The third and final pathway is reconstructive relationalism, where contractual doctrines are remodelled,[32] as illustrated by the Canadian judgment of *Bhasin*.[33]

Each of Tan's pathways show that the notion of the relational contract makes it clear that the terms of the agreement are only part of the equation. This point is also shared by the contextualism movement. According to Hugh Collins, there are three levels of social relations that shape contracts.[34] First, the written contract represents the frame of reference. Second, the economic relations illustrate the rational self-interest of the parties. Third, trust impacts on every social interaction between the parties. Economic relations and trust form the core of the implicit dimensions of a contract. The cement between these elements is the legitimate expectations of the parties,[35] meaning an expectation that a benefit or right will be obtained as the contract is performed. For example, parties rely on the good faith of the other.

The notion of good faith is well known to civil law lawyers. Teubner famously describes good faith as a legal irritant, whose implementation in English law through European Union directives would start a domino effect in contract law that would 'irritate British legal culture considerably'[36] and also 'trigger deep, long-term changes from highly formal rule-focused decision-making in contract law towards a more

28 Sandrine Tisseyre, *Le rôle de la bonne foi en droit des contrats – Essai d'analyse à la lumière du droit anglais et du droit européen* (LGDJ 2012) 280.

29 Zhong Xing Tan, 'Disrupting doctrine? Revisiting the doctrinal impact of relational contract theory' (2019) 39 Legal Studies 98, 105.

30 Ibid 107.

31 Ibid 108; *Yam Seng Pte Limited v International Trade Corporation Limited* [2013] EWHC 111 (QB) [142].

32 Ibid 111.

33 *Bhasin v Hrynew* [2014] 3 SCR 494.

34 See Hugh Collins, 'Discretionary powers in contracts' in David Campbell, Hugh Collins and John Wightman (eds), *Implicit Dimensions of Contract: Discrete, Relational, and Network Contracts* (Hart 2003) 250.

35 Ibid.

36 Gunther Teubner, 'Legal irritants: good faith in British law or how unifying law ends up in new divergences' (1998) 61 Modern Law Review 11, 20.

discretionary principle-based judicial reasoning'.[37] Interestingly, in France, good faith and relational contracts have been analysed. For instance, Busseuil developed a definition of the relational contract using two main pillars: the legal link between the contractual obligation and the relationship and the notion of *favor contractus*.[38] Parties do not mind owing each other favours[39] if this means that the contractual relationship is maintained and even flourishes: the contractual link between the parties will prevail. From this perspective, Busseuil considers that good faith has an important role to play to ensure the contract is adapted to comply with a new set of circumstances or to remedy the absence of some terms in the contract. While good faith is a principle of contract law well known to civil lawyers, its place if any in common law jurisdictions is less certain.[40]

Community standards, morals and contract law

No matter the jurisdiction, reasons for entering the relationship can be implied by standards from a specific industry, the broader business context, and the general community or even the idealised general community.[41] This is also very close to social relationship theory which emphasises the importance of the societal context, and the reality faced and understood by the community, in the shaping of contract law. This approach 'confines expectations and is imbedded in conventions, norms, mutual assumptions and unarticulated expectations'.[42]

Laws and morals are traditionally two different notions.[43] According to an old English legal doctrine, 'From a dishonourable cause, an action

37 Ibid 21.

38 Guillaume Busseuil, *Contribution à L'étude De La Notion De Contrat En Droit Privé Européen* (LGDJ 2008) 350.

39 Jan B Heide and George John, 'Do norms matter in marketing relationships?' (1992) 56(2) Journal of Marketing 32, 39.

40 Teubner (n 36 above); Wayne Courtney, 'Good faith and termination: the English and Australian experience' (2019) 1 Journal of Commonwealth Law 185; Martin Hogg, 'The implication of terms-in-fact: good faith, contextualism and interpretation' (2018) 85(6) George Washington Law Review 1660; Ewan McKendrick, 'Good faith in the performance of a contract in English law' in Larry DiMatteo and Martin Hogg (eds), *Comparative Contract Law: British and American Perspectives* (Oxford University Press 2015) 196; recent cases in Canada have been slowly canvassing an organising principle of good faith ever since *Bhasin v Hrynew* [2014] SCC 71 and more recently *Wastech Services Ltd v Greater Vancouver Sewerage and Drainage District* [2021] SCC 7.

41 Jeannie Paterson, 'The standard of good faith performance: reasonable expectations or community standards?' in Michael Bryan (ed), *Private Law in Theory and Practice* (Routledge-Cavendish 2007) 158.

42 Hugh Collins, 'Introduction' in Campbell et al (eds) (n 34 above) 2.

43 Beatrice Jaluzot, *La Bonne Foi dans les Contrats: Etude Comparative des Droits Francais, Allemand et Japonais* (Dalloz 2001) 62.

does not arise.'[44] This means that there is a need for a breach of law for a remedy to be available; a breach of morals is not sufficient. The word 'morals' is said to have originated from the Latin *moralis*, coined by Cicero.[45] Morals vary from one person to another and are mostly imposed by individuals upon themselves.[46] There is 'a fundamental difference between the law that expresses a moral principle and that law that is only a social regulation'.[47] Morals and law have different aims: on the one hand, morals are linked to the individual, whereas law is addressed to society. They have different sources: morals come from the conscience of the individual and law is imposed by external institutions. Finally, morals are broader and vaguer, while law enacts precise rules. As Devlin explains: 'Legalisation is seen as the natural enemy of morality, for morality is at its best when each case is judged entirely on its merits.'[48]

To provide context to the debate surrounding the understanding of the notion of good faith in contracts, it is important to reflect on the origins of the distinction between law and morals to demonstrate that they have always been intertwined. Aristotle did not use the notion of good faith but identified three virtues: liberality, fidelity and commutative justice. Firstly, Aristotle understood liberality as using resources sensibly.[49] This meant, for instance, that people should not be prodigal, that is, waste their substance (their money).[50] Secondly, breaking a promise was being unfaithful to one's word.[51] An example of the second virtue can be found in article 1104 of the French Civil Code. The renewed attention given to this French Civil Code provision after the Second World War[52] and the increasing impact of morality on different relationships, including contractual relations, slowly led to the development of new obligations and duties of parties, such as the duties to disclose information and to cooperate in the negotiation and

---

44 *Ex turpi causa non oritur actio.*

45 *Chambers Dictionary of Etymology* (Chambers 2008).

46 Jaluzot (n 43 above) 61.

47 Patrick Devlin, *The Enforcement of Morals* (Oxford University Press 1965) 60.

48 Ibid 46.

49 James Gordley, 'Some perennial problems' in James Gordley (ed), *The Enforceability of Promises in European Contract Law* (Cambridge University Press 2001) 4.

50 Aristotle, *Nicomachean Ethics* (first published 350 BC) book IV.

51 Gordley, 'Some perennial problems' (n 49 above) 6. Primary source: Cajetan, *Commentaria to Thomas d'Aquinas Summa Theologiae* (Padua 1698) pt II-II, q 88, art 3; q 110.

52 Philippe Le Tourneau and Matthieu Poumarède, 'Bonne foi' (2009) Repertoire Civil nn 3–4, n 18.

performance of the agreement.[53] The developments have contributed to good faith being applicable to all stages of the contract in article 1104 following reform in 2016. These obligations all mirror the Aristotelian virtues of fidelity and commutative justice, the latter being associated with equitable fairness.[54] These ideas are also echoed in the common law commentary. For instance, Andrew Gold discusses the relationship between morality and loyalty:

> Promises might also ground a morally significant loyalty obligation, depending on one's theory of loyalty. For example, there may be cases in which an individual promises to be loyal, thus creating a moral duty to be loyal in light of that promise. Accordingly, even if loyalty lacks a moral basis as a general matter, in specific contexts loyalty can take on a moral dimension – loyalty and morality are at least sometimes linked.[55]

The idea that promises should be kept was also shared by Thomas Aquinas,[56] in the same way that Aristotle asked parties in a contractual relationship to keep to their word.[57] Thirdly, the idea of commutative justice or the will to exchange resources of equivalent value, so that neither party is enriched at the expense of the other, was also reinforced by Thomas Aquinas.[58] Therefore, the three Aristotelian virtues were maintained during the medieval period.[59] Reflecting this philosophy, contracts were classified under two broad categories: liberalities or donations and commutative justice,[60] referring to contracts as the Romans understood them, that is, consensus contracts.[61]

The understanding of the doctrine of good faith was revived by the works of Baldus, a leading medieval Roman law scholar of the fourteenth century in Italy. Before then, good faith was understood as consisting

53 Muriel Fabre-Magnan, *Droit des Obligations* 2nd edn (PUF 2010) 63; Rémy Cabrillac, *Droit Européen Comparé des Contrats* (LGDJ 2012) 3; Busseuil (n 38 above) 552.

54 James Gordley, 'Good faith in contract law in the medieval *ius commune*' in Reinhard Zimmermann and Simon Whittaker (eds), *Good Faith in European Contract Law* (Cambridge University Press 2000) 107; Anton-Hermann Chroust and David L Osborn, 'Aristotle's conception of justice' (1942) 17 Notre Dame Law Review 129, 136.

55 Andrew S Gold, 'Accommodating loyalty' in Paul B Miller and Andrew S Gold (eds), *Contract, Status, and Fiduciary Law* (Oxford University Press 2016) 5.

56 Gordley, 'Some perennial problems' (n 49 above) 4. Primary source: Thomas Aquinas, *Summa Theologiae* pt II-II, q 88, art 3, ad 1; q 110, art 3, ad 5.

57 Aquinas (n 56 above). Primary source: Cajetan, *Nicomachean Ethics* book IV, ch vii, 1127a–1127b.

58 Aquinas (n 56 above) Q 61, art 3.

59 Ibid.

60 Ibid art 5.

61 Cause and consideration may be different but seek to achieve the same goal: the promise is binding and has a reason to exist.

of three obligations:[62] to keep to one's word, to not take advantage by misleading or driving too harsh to a bargain,[63] and to abide by obligations an honest person would recognise.[64] The first obligation has always been part of the development of good faith and led to *pacta sunt servanda*.[65] Baldus revived the works of Thomas Aquinas and Aristotle to arrive at a better understanding of the notion of good faith. He understood the concept as an obligation not to become enriched at the expense of the other party.[66]

This echoes some of the recognition that can be found in Australia. For instance, there is a difference between acting in the interests of another and taking their interests into consideration. The concept of good faith only applies in the latter situation and must be differentiated from fiduciary duties.[67] Fiduciary duties are equitable duties that can exist together with contractual duties.[68] Fiduciaries must not profit from their position and must avoid and disclose conflicts of interest. Remedies may be restitution (disgorging profits) and compensation if a breach results in loss to the beneficiary. The express fiduciary duty is to act in the interest of another. In *McKenzie v McDonald*,[69] this duty was defined as one party having powers and discretions that affect the interests of the other, the latter putting trust and confidence in the actions of the former. The category is open-ended and can overlap with other doctrines such as unconscionability.

This analysis shows that the traditional division between law and morals does not consider the necessary convergence between the two notions. While they have different characteristics, laws should not go against morals and some laws find their source in morals.[70] As Rowan explains, 'depending on the context, the right-holder can be required to refrain from acting dishonestly, outside the limits of the

62 Gordley (n 54 above) 94.

63 Ibid 99–101.

64 See Reinhard Zimmermann, *The Law of Obligations: Roman Foundations of the Civilian Tradition* (Oxford University Press 1990) 664–671.

65 See eg Ulpian, *Digest* 2, 14, 77; for a longer discussion see also Alexis Keller, 'Debating cooperation in Europe from Grotius to Adam Smith' in William Zartman and Saadia Touval (eds), *International Cooperation: Extents and Limitations of Multilateralism* (Cambridge University Press 2010) 19.

66 Baldus de Ubaldis, *Consilia, Sive Responsa* (1575); Gordley (n 54 above) 93, 94.

67 Paul Finn, 'The fiduciary principle' in T G Youdan (ed), *Equity, Fiduciary and Trusts* (Carswell 1989) 4, cited in Andrew Terry and Cary Di Lernia, 'Franchising and the quest for the holy grail: good faith or good intentions' (2009) 33 Melbourne University Law Review 542, 554.

68 J W Carter and M P Furmston, 'Good faith and fairness in the negotiation of contracts part I' (1994) 8 Journal of Contract Law 1, 6; Finn (n 67 above) 4, cited in Terry and Di Lernia (n 67 above) 554.

69 [1927] VLR 134 (Dixon J).

70 Jaluzot (n 43 above) 63.

rights, inconsistently with the purpose for which it was conferred or without a legitimate interest or any proper basis'.[71] The concept of good faith is an example of this convergence. In France it has moved from moral rule to legal norm: 'good faith is one of the means used by the legislature and the courts to allow the moral rule to penetrate in law'.[72] The doctrine of good faith protects the legitimate expectations of the parties and ensures both procedural and substantive fairness in contractual dealing. It regulates behaviours[73] and, while the intention of the parties is interpreted objectively, the notion of legal expectation is what comes closest to the theory of subjective rights as it is known in civil law.[74] In common law countries, the moral view of contract law is that the good person should deal fairly.[75] The pragmatic view is that keeping faith does not matter.[76] But this is a contrarian view in the common law. The doctrine of good faith faces opposition with opponents insisting that it is difficult to decide where morals stop and where law begins.[77] Yet, the first part of this article has shown how actions based on good faith can be used to broaden justice and to punish fraudulent behaviour, including misleading or taking advantage of the other party.[78] *Carter v Boehm*, a landmark case in insurance contract law, led to the recognition of the doctrine of good faith in UK insurance contracts. It demonstrates the importance of the concept through the duty of disclosure: 'the reason of the rule which obliges parties to disclosure, is to prevent fraud, and to encourage good faith'.[79] This shows that good faith is also about fidelity to the bargain and that such fidelity is required and encouraged in some transactions including insurance contracts.

Good faith and relational contract: two sides of one coin

The discussion above illustrates that one underpinning of the spectrum of contracts is the influence of moral values on contracting.

71 Rowan (n 17 above) 13, at 1068; See *Braganza v BP Shipping Ltd* [2015] UKSC 17 (SC); *Renard Constructions (ME) Pty Ltd v Minister for Public Works* (1992) 26 NSWLR 234; *Garry Rogers Motors (Aust) P/L v Subaru (Aust) P/L* (1999) 21 ATPR 41-703.

72 Jaluzot (n 43 above) 65. Georges Ripert, *La Règle Morale Dans Les Obligations Civiles* (LGDJ 1949) 157: 'la bonne foi est l'un des moyens utilisés par le législateur et les tribunaux pour faire pénétrer la règle morale dans le droit'.

73 Jaluzot (n 43 above) 66.

74 Busseuil (n 38 above) 587.

75 Devlin (n 47 above) 43.

76 This is where an efficient breach occurs.

77 R Goode, 'The concept of "good faith" in English law' (Centro di Studi e Richerche di Diritto Comparato e Straniero, Saggi, Conferenze e Siminari 2, Rome 1992).

78 Gordley (n 54 above) 100.

79 *Carter v Boehm* (1766) 3 Burr 1905, 1162, 1165.

This is further demonstrated by the idea that any 'exchange behaviour includes a combination of the ten common contract norms, containing at a minimum solidarity and reciprocity'.[80] Therefore, the cooperative nature of the exchange is seen as a 'given' in relational contract theory if the contract is deemed relational. Cooperation itself is a well-known contract law principle. Understanding the importance given to the contractual promise and the moral limits to party autonomy provides valuable insights into the understanding of good faith in contract law.[81] But beyond this, solidarity keeps creeping into contractual exchanges, meaning that '[p]arties are committed to improvements that may benefit the relationship as a whole, and not only the individual parties'.[82] This reasoning is consistent with the idea that parties need to take into consideration the interests of the other party. Good faith is presented as an economic expectation of the parties, whereby cooperative norms actually lower costs.[83]

Good faith does not require altruism, or subjugation of self-interest.[84] But breaking a promise goes against the notion of parties sticking to their bargain: 'The aim of contract law ... is to make things better.'[85] In principle, parties to a contract enter into the contract to see it performed.[86] If we accept that the promise is at the core of the theory of contractual obligation then, according to Fried, there is a moral obligation to make a promise binding.[87] Fried argues that good faith requires loyalty to the promise and that contract law imposes legal obligations that are convergent with moral obligations.[88] Reciprocity and contractual solidarity bring ideas of fairness and justice into contract law. Good faith is seen as the moral and legal obligation to ensure parties cooperate. Not only is anti-cooperative behaviour discouraged,[89] but parties are encouraged to cooperate. This has led to the judicial recognition of some principles of fairness in contracting.

---

80 Chapi F Cimino, 'The relational economics of commercial contracts' (2015) 3 Texas A&M Law Review 91, 100.

81 Finn (n 18 above) 149.

82 Ibid.

83 Cimino (n 80 above) 114.

84 Liam Brown, 'The impact of section 51AC of the TPA on commercial certainty' (2004) 28 Melbourne University Law Review 589, 605.

85 Liam B Murphy, 'The practice of promise and contract' in Gregory Klass, George Letsas and Prince Saprai (eds), *Philosophical Foundations of Contract Law* (Oxford University Press 2014) 153.

86 See Ibid.

87 Charles Fried, *Contract as Promise* (Oxford University Press 2015) 146; see also Hillman (n 20 above) 12.

88 Fried (n 87 above) 147.

89 Ian Macneil, 'Efficient breach of contract: circles in the sky' (1982) 68 Virginia Law Review 947, 968.

A CHRONOLOGY OF JUDICIAL DEVELOPMENTS: TWO STEPS FORWARD, ONE STEP BACKWARD?

While the reasonable exercise of discretionary rights and the notions of cooperation and collaboration have been the subject of many court decisions which examined the behaviour of contractual parties in the performance of their contract, the concept of the relational contract is only emerging.

Pre *Bates*: the hesitancy of the courts

While the notion of relational contract has been known since Macneil's seminal piece, it has not been used in judicial decisions in Australia to the same extent it has recently been in the UK. In *Johnson v Unisys Ltd*,[90] Lord Steyn suggested the contract of employment could be described as a relational contract.[91] In the antipodes, Finn J in *GEC Marconi Systems* articulated that a relational contract is a contract which 'involves not merely an exchange, but also a relationship, between the contracting parties'.[92]

It is the UK decision of *Yam Seng*[93] that truly reignited the possibility that some long-term contractual relationships 'between parties who make a substantial commitment' have a special status:

> While it seems unlikely that any duty to disclose information in performance of the contract would be implied where the contract involves a simple exchange, many contracts do not fit this model and involve a longer term relationship between the parties which they make a substantial commitment. Such 'relational' contracts, as they are sometimes called, may require a high degree of communication, cooperation and predictable performance based on mutual trust and confidence and involve expectations of loyalty which are not legislated for in the express terms of the contract but are implicit in the parties' understanding and necessary to give business efficacy to the arrangements.[94]

---

90 [2003] 1 AC 518. See Matthew Boyle, 'The relational principle of trust and confidence (2007) 27(4) Oxford Journal of Legal Studies 633; Hugh Collins, 'Is a relational contract a legal concept?' in S Degeling, J Edelman and J Goudkamp (eds), *Contracts in Commercial Law* (Thomson Reuters 2016) 37; Gabrielle Golding, 'Employment as a relational contract and the impact on remedies for breach' (2021) 30(2) Griffith Law Review 270; Douglas Brodie, 'Relational contracts' in M Freeland et al (eds), *The Contract of Employment* (Oxford University Press 2016) 145.

91 *Johnson v Unisys Ltd* [2003] 1 AC 518 at 532 [20].

92 *GEC Marconi Systems Pty Ltd v BHP Information Technology Pty Ltd* [2003] FCA 50; (2003) 128 FCR 1 [224].

93 *Yam Seng PTE v International Trade Corp Ltd* [2013] EWHC 111 (QB) [131], [142], [145].

94 Ibid [142].

This definition was followed in *D&G Cars* and led to parties' duty to act honestly and with integrity in executing the contract.[95]

In the 2018 decision of *Sheikh Al Nehayan v Kent*,[96] LJ Leggatt emphasised that relational contracts are about the commitment of the parties to collaborate, supported by the idea that parties will 'act with integrity and in a spirit of cooperation. The legitimate expectations which the law should protect in relationships of this kind are embodied in the normative standard of good faith.'[97] The same year, the concluding remarks of LJ Jackson in *Amey Birmingham Highways Ltd v Birmingham City Council* also characterised the contract in dispute as a relational contract but did not venture into developing this more broadly, fearing to engage in further 'contentious considerations'.[98]

The notion of the relational contract was also more recently mentioned and discussed in Australia in Commonwealth *Bank of Australia v Barker*.[99] In their joint judgment, Chief Justice French, Justice Bell and Justice Keane referred to Lord Steyn's judgment in *Johnson v Unisys Ltd*.[100] This case also provided relevant *dicta* on good faith, although this legal principle had not been argued by the parties themselves and the court did not have to decide on the application of good faith in that situation.

While good faith has been discussed in Australia, the notion of relational contracts has rarely been mentioned in Australian case law. Recently, parties have argued that the agreement at the heart of the legal dispute was a relational contract, where the contract 'involves not merely an exchange but also a relationship between the contracting parties',[101] quoting *GEC Marconi*. In *Binaray Pty Ltd v RAMS*, the designation of a franchise agreement as relational did not import any special rules of construction that would not otherwise apply.[102] There are also other topical illustrations of the use of relational contracting. The notion of relational contracting is also surfacing in defence contracts, although the data on this phenomenon is limited.[103] Beyond the notion of the relational contract itself, Australian courts

95 *D&G Cars Ltd v Essex Police Authority* [2015] EWHC 226 (QB) [174]–[176].
96 [2018] EWCH 333 (Comm).
97 Ibid [167].
98 [2018] EWCA Civ 264 [92].
99 *Commonwealth Bank of Australia v Barker* [2014] HCA 32.
100 Ibid [17].
101 *Centreplex Pty Ltd v Noahs Rosehill Waters Pty Ltd* [2019] WASC 252 [102].
102 *Binaray Pty Ltd (ACN 119 724 211) as Trustee for the Allen Family Trust v RAMS Financial Group Pty Limited* (ACN 105 207 538) [2019] QSC 33 [92].
103 Bruce McLennan, 'How Australia's Department of Defence harnessed the power of the relationship' (16 August 2018) Contracting Excellence Journal; Kate Vitasek, 'Relational contracting on the rise with the success of the Australian navy' (*Forbes*, 30 November 2016).

have enforced parties' duty to cooperate,[104] and the duty not to exercise a discretionary contractual right arbitrarily,[105] capriciously or for an external purpose.[106] Most of the disputes have related to performance and termination, two aspects of the life of the contract that are interlinked.[107] Beyond this, some Australian courts have enforced a duty on parties to act in good faith,[108] although the duty can still be excluded through contract drafting.[109] Meanwhile in the UK, the Supreme Court decision in *Bates v Post Office* explicitly brought good faith and relational contract together. The confident decision by Fraser J merits closer attention.

Bates or the explicit linking of good faith and relational contracting

Facts, context and relationship

Bates v Post Office dealt with 550 claimants, including Bates, who were responsible for running Post Office branches. Most of the claimants were sub-postmasters, but some were also employees. The difference matters as each group had different contractual terms. The March 2019 decision focuses on the former group. The years 1999 and 2000 saw the rollout of a new electronic point of sale and accounting system using software called Horizon. The Post Office made it mandatory for the claimants to use the system.[110] There was, however, an issue with the way the system operated. The claimants argued that the new system contained many coding errors, leading to discrepancies in branches' accounting and ultimately shortfalls, first of hundreds of pounds and then of thousands within months.[111] Some claimants went broke, some were locked out of their Post Offices, and some contracts were terminated abruptly. The claimants claimed the Post Office's

104 *Butt v McDonald* (1896) 7 QLJ 68.

105 *Renard Constructions (ME) Pty Ltd v Minister for Public Works* (1992) 26 NSWLR 234.

106 *Garry Rogers Motors (Aust) P/L v Subaru (Aust) P/L* (1999) 21 ATPR 41-703.

107 Courtney (n 40 above).

108 *Renard Constructions (ME) Pty Ltd v Minister for Public Works* (1992) 26 NSWLR 234, 246; *Sigiriya Capital Pty Ltd v Scanlon* [2013] NSWCA 401; *Burger King Corporation v Hungry Jack's Pty Ltd* (2001) 69 NSWLR 558; *Bundanoon Sandstone Pty Ltd v Cenric Group Pty Ltd* [2019] NSWCA 87.

109 *Vodafone Pacific Ltd v Mobile Innovations Ltd* [2004] NSWCA 15; *Growthbuilt Pty Ltd v Modern Touch Marble & Granite Pty Ltd* (2021) [2021] NSWSC 290 [60]–[62].

110 There was no opt-out option, although, as Fraser J stated, it is understandable for a large company not to take a piecemeal approach: *Bates v Post Office Ltd (No 3)* [2019] EWHC 606 (QB) [7].

111 No claims were brought against ICL or Fujitsu who then owned the software.

actions were harsh and unfair.[112] There were numerous claims and cases dealing with this dispute, including damages for financial loss, personal injury, deceit, duress, unconscionable dealing, harassment and unjust enrichment as well as criminal actions all brought against the Post Office.[113] The March decision dealt with the nature of the contract and the terms it contained.

Limitation clauses and burden of proof

The first issue to be determined was whether the claimants had to pay the Post Office in cases of financial shortfalls. Fraser J made it clear this was not about determining the existence of these shortfalls. Fraser J would later find that there had been shortfalls and that these were due to a system that was not robust enough.[114] But first, since the Post Office had demanded payment and taken drastic action, the judge looked into the agreement. As with any written agreement, the first issue was to examine the express provisions in the contract itself. After re-emphasising the importance of objective interpretation, the judge determined that the contract was very detailed but that the sub-postmasters had no control over the terms and their negotiation.[115]

The *contra proferentem* rule, which holds that a term should be interpreted against the one who is arguing it, is no longer applied in the UK.[116] Fraser J considered the rule a 'historical remnant',[117] instead preferring to rely on the natural meaning of the words. He quoted Lord Neuberger MR's judgment of 2011, stating: 'the words used, commercial sense and the documentary and factual context, are and should be normally enough to determine the meaning of a contractual provision'.[118] This was then echoed in *Arnold v Britton*[119] and *Persimmon Homes Ltd v Ove Arup & Partners Ltd*[120] in which Jackson LJ determined that in commercial contracts where there is equal bargaining power the *contra proferentem* rule has a limited role. However, as Fraser J rightly pointed out, the sub-postmasters were not able to negotiate the contract. Leaning on the natural meaning of the words,[121] the court considered that, under the contract, the sub-

112 *Bates v Post Office Ltd (No 3)* [2019] EWHC 606 (QB) [20].
113 Ibid [10].
114 Ibid.
115 Ibid [638].
116 Ibid [638], [653].
117 Ibid [635].
118 *K/S Victoria Street v House of Frazer (Stores Management) Ltd* [2011] EWCA Civ 904 [63].
119 [2015] UKSC 36.
120 [2017] EWCA 373.
121 *Bates v Post Office Ltd* (No 3) [2019] EWHC 606 (QB) [646].

postmasters would only be liable financially if the loss was due to their own negligence, carelessness or error and the Post Office had to prove that the loss fell in that category.[122] The newer contract contained a much broader limitation clause according to which the sub-postmasters would be liable for any loss unless it was due to criminal acts the sub-postmasters could not have prevented.[123] However, the Post Office did not demonstrate an actual and real loss, meaning a loss and resulting shortfall – only a loss according to Horizon data.[124] Furthermore, the Post Office did not satisfy the reasonableness test under the Unfair Contract Terms Act 1977 (UK).[125]

Recognising the contract as a relational agreement

The question of whether the contracts in the dispute were relational contracts was one of the most important issues of the litigation and judgment.[126] Would such a determination lead to the implication of particular terms in the contract? Would these include the 21 different terms laid out by the claimants before the court?[127] To what contractual powers, discretions and/or functions do such terms apply? For Fraser J, taking a contextual approach was key to deciding the case.

The claimants claimed that:

> The [sub-postmaster] contracts were replete with power and discretion in the hands of the Defendant. In all the circumstances, they included an implied term of trust and confidence and/or were relational contracts imposing obligations of good faith on the Defendant (including duties of fair dealing and transparency, trust and confidence and co-operation). There were also implied terms, including obligations on the Defendant: not to act in an arbitrary, irrational or capricious manner in decision making affecting the Claimants; to provide adequate training and support to the Claimants (particularly if and when it imposed new working practices or systems or required the provision of new services); properly to execute all transactions which the Claimants effected; properly to account for, record and explain all transactions and any alleged shortfalls which were attributed to the Claimants; and properly and fairly to investigate any such alleged shortfalls.[128]

The parties agreed to the implication of two implied terms.[129] The first implied term required each party not to take any step to inhibit or prevent the other party from complying with its obligations under

122 Ibid.
123 Ibid [682].
124 Ibid [687].
125 Ibid [1108]–[1110].
126 Ibid [31].
127 Ibid [45].
128 Ibid [326].
129 Ibid [698].

or by virtue of the contract. The second term required each party to provide the other with such reasonable cooperation as necessary to the performance of the other's obligations under or by virtue of the contract. These are non-contentious provisions that are already part of contract law in the UK.[130]

The issue was that the claimant then asked for another 21 terms. The Post Office considered that these were too many to be implied. The judge made sure not to take into consideration hindsight when determining whether the terms would be implied.[131] The court focused on business efficacy and

> what notional reasonable people, in the position of the parties at the time at which they had been contracting, would have agreed and that it was a necessary but not sufficient condition for implying a term that it appeared fair or that the court considered that the parties would have agreed it if it had been suggested to them.[132]

This is in line with the decision in *Marks and Spencer plc v BNP Paribas*.[133] The judge ultimately recognised 17 terms that were implied into the agreement. However, to reach this conclusion, the judge first considered the nature of the contract, namely its relational characteristic, as the implications of these terms depended on it.

The Post Office argued there was no such type of contract. It relied upon *Chitty on Contract* to justify its position and restrict good faith to honesty. But Fraser J disagreed and considered that good faith was more than honesty. He used the judgment of Dove J in *D&G Cars Ltd v Essex Police Authority* and referred to integrity and the need to maintain 'the mutual trust and confidence between the parties in this long-term relationship without necessarily amounting to the telling of lies, stealing or other definitive examples of dishonest behaviour'.[134] After acknowledging the persuasive effect of academic learning in the common law,[135] the judge disagreed[136] with Chitty's position that 'the implication of such an implied term applicable generally (or even widely) to commercial contracts would undermine to an unjustified extent English law's general position rejecting a general legal requirement of good faith'.[137] The judge used case law to support this determination.[138]

130 *Mackay v Dick* (1881) 6 App Cas 251; *Stirling v Maitland* [1864] 122 ER 1043.
131 *Bates v Post Office Ltd (No 3)* [2019] EWHC 606 (QB) [745].
132 Ibid [694].
133 [2015] UKSC 72.
134 *D&G Cars Ltd v Essex Police Authority* [2015] EWHC 226 (QB) (Dove J) [175].
135 *Bates v Post Office Ltd (No 3)* [2019] EWHC 606 (QB) [709].
136 Ibid [710]–[711].
137 Ibid [708].
138 Ibid [712]–[721].

Fraser J first acknowledged that the lack of equal standing of the parties is not what makes a relational contract, before using *Yam Seng* as a springboard to lay out a set of criteria to determine what makes a contract relational.[139]

> 1. There must be no specific express terms in the contract that prevents a duty of good faith being implied into the contract.
>
> 2. The contract will be a long-term one, with the mutual intention of the parties being that there will be a long-term relationship.
>
> 3. The parties must intend that their respective roles be performed with integrity, and with fidelity to their bargain.
>
> 4. The parties will be committed to collaborating with one another in the performance of the contract.
>
> 5. The spirits and objectives of their venture may not be capable of being expressed exhaustively in a written contract.
>
> 6. They will each repose trust and confidence in one another, but of a different kind to that involved in fiduciary relationships.
>
> 7. The contract in question will involve a high degree of communication, co-operation and predictable performance based on mutual trust and confidence, and expectations of loyalty.
>
> 8. There may be a degree of significant investment by one party (or both) in the venture. This significant investment may be, in some cases, more accurately described as substantial financial commitment.
>
> 9. Exclusivity of the relationship may also be present.[140]

Fraser J stated that the list was not definitive and that no criterion was determinative except for the one that there should not be an express provision to prevent the duty of good faith being implied.[141] By considering more than the written agreement between the parties, Fraser J adopted a position akin to the contextualists who base their theory on the fact that there is more than just an agreement between the parties.[142] The judgment also shows that a contractual situation is made up of a written agreement as well as some implicit understandings. These understandings need to be considered to ensure that the intentions of the parties, and their reasons for entering the relationship, are clearly reflected. According to Macaulay, the concept

139 Ibid [722].
140 Ibid [725].
141 Ibid [726].
142 See eg Campbell et al (eds) (n 34 above).

of the relational contract can be used in different situations: either as a way to encourage settlement between parties and to interpret indeterminate legal principles; or to reduce the costs associated with a long-term relationship due to the lack of foreseeability associated with it.[143] *Bates* illustrates the latter point as the sub-postmasters could not have foreseen the discrepancies that resulted from the rollout of the Horizon software.

Furthermore, the relationship between the parties was more than 'purely commercial' because the Post Office is required by the government to maintain a broad network of branches across the country, even in locations that are not viable. The role of the Post Office as a public service to the community at large was therefore taken into consideration, and justifiably so. The notion of trust between the Post Office and the sub-postmasters is essential and this was submitted by both parties.[144] Trust was also seen as paramount in the Post Office-related activities carried out by the sub-postmasters and members of the public wishing to use that Post Office branch. Not only did the benefits provided under the sub-postmasters' contracts have similarities with an employment relationship,[145] including the entitlement to holiday substitution allowance, but it required from the sub-postmasters a major degree of investment and significant personal financial commitment to running that branch,[146] with the Post Office conducting thorough checks on the applicants before they became sub-postmasters. A relational contract does not need an imbalance in power between the parties, although there clearly was one in this instance.[147] Other non-essential features of the relationship that were taken into consideration included the residential accommodation in which the sub-postmasters themselves (and potentially other family

143 See Stewart Macaulay, 'The real and the paper deal' in Campbell et al (eds) (n 34 above) 83.

144 *Bates v Post Office Ltd (No 3)* [2019] EWHC 606 (QB) [728]. In the UK, trust and confidence are implied as a matter of law in employment contracts, see *Malik and Mahmud v Bank of Credit and Commerce International SA* [1997] UKHL 23. This has been rejected in *Australia in Commonwealth Bank of Australia v Barker* (2014) 253 CLR 169.

145 As noted at n 144 above, the notions of trust and confidence are implied as a matter of law in all employment contracts.

146 'The Post Office knew not only of the size of this investment, but the source of an incoming SPM's funds, as these were included in the business plans submitted by the SPMs. If the source of funds was not identified, this information would be sought by the Post Office.' *Bates v Post Office Ltd (No 3)* [2019] EWHC 606 (QB) [728].

147 Ibid [724].

members) lived,[148] and the fact that the Post Office shares features with a public body.[149]

Bates is an important decision because it further acknowledges the notion of relational contract as an important part of contract law and not, as Collins suggested, a 'passing fad'.[150] Furthermore, and despite being a case of first instance, it is the first judicial decision that provides a list of clear criteria for a relational contract in law. By considering the duration and the context of the contractual relationship, decisions such as *Yam Seng* and *Bates* add a new dimension to contract law doctrine. It is not the type of party that matters, such as tenant–landlord, consumer–business, employer–employee,[151] but the relationship itself.

For Fraser J, the formula to determine whether a contract is relational consists of three elements: one must take into consideration the relations between the parties, the terms of the contract and the context of the transaction. This will determine whether the contract is a relational one.[152] Ultimately, the judgment provides some guidance on the notion of the relational contract, but also blurs the boundaries by stating that the list of criteria is not exhaustive and that just because the contract is long term or, for instance, a franchise does not mean it is in fact relational. This leaves many pondering the broader implications if any for contract law. One clear implication of a contract being deemed relational is that it is likely that a duty to perform in good faith will be implied in law into the agreement.

Implying good faith

Bates illustrates that doctrines such as cooperation and good faith are included in the implicit dimensions of contracts.[153] The court found that the contract did not contain terms that excluded good faith. Since the contract was held to be relational in nature, the terms included an implied obligation of good faith.

Fraser J determined that there

> is a specie of contracts, which are most usefully termed 'relational contracts', in which there is implied an obligation of good faith (which is also termed 'fair dealing' in some of the cases). This means that the parties must refrain from conduct which in the relevant context would be regarded as commercially unacceptable by reasonable and honest

148 Ibid [729].
149 Ibid [730].
150 Ibid [736] with a reference to Collins (n 90 above) 38.
151 Gabrielle Golding, 'Employment as a relational contract and the impact on remedies for breach' (2021) 30(2) Griffith Law Review 270.
152 *Bates v Post Office Ltd (No 3)* [2019] EWHC 606 (QB) [721].
153 Macaulay (n 143 above) 52.

> people. An implied duty of good faith does not mean solely that the parties must be honest, but also '[t]ransparency, co-operation, and trust and confidence'.[154]

Once the contracts formed between the Post Office and sub-postmasters were designated as relational contracts, then there was an implied duty of good faith in the agreements. That implied duty of good faith applies to both parties to the contract.[155] In this instance, however, it was the conduct of the Post Office that was found to breach good faith.

> The Post Office is not therefore entitled to rely upon the Branch Trading Statements, for any period in respect of which a SPM notified a dispute to the Helpline, as a settled account between agent and principal. ... Nor do SPMs bear the burden of demonstrating that the Branch Trading Statement is wrong for such a period.[156]

There is no restriction upon the Post Office in terms of how this discretion can be exercised, other than that the discretion available to the Post Office should be exercised for a proper purpose and in accordance with the implied duty of good faith. The fact that the case extensively deals with the implied duty to act in good faith further advances the reform of contract law, even though its recognition is not (yet?) a *fait accompli*. While the concept of good faith is controversial, the notion that a party to the contract should behave decently when performing its contractual obligations is less of a burning issue. But, if good faith means more than cooperation, and cooperation is one of the only parts of good faith to be enforced before the courts, why would parties act in good faith? The need for consequences is driving the slowness of the recognition of good faith in some long-term contracts. This hesitancy is further demonstrated by more recent decisions.

Post *Bates*: the reluctance on the qualification of relational contracts remains

On 12 November 2019, Lord Coulson rejected the appeal from the Post Office on the grounds there was little realistic prospect of success. He highlighted the conduct of the Post Office during the trial and agreed with Justice Fraser that the contract was relational and that a duty of good faith was applicable, after analysing the context of the relationship between the Post Office and the sub-postmasters.[157] On 11 December 2019, the Post Office and the claimants released a joint statement terminating the litigation after the parties agreed to

---

154 *Bates v Post Office Ltd (No 3)* [2019] EWHC 606 (QB) [711].

155 Ibid [1113].

156 Ibid [1116].

157 *Bates v Post Office* (Judgment on PTA, 22 November 2019, A1/2019/1387/PTA) (Coulson LJ).

resolve the dispute through mediation.[158] A few days later, the court released its judgment on the Horizon issues.[159] It decided that there had been bugs in the system.[160] Changes to the system since 2010,[161] the settlement out of court and the fact that the sub-postmasters were cleared of corrupting data in early 2021 show that the case is now closed. But will the March decision make an impact on the further development of the relational contract in UK contract law? Recent cases show that identifying a contract as relational is still a difficult process, and courts are hesitant to fully engage with it. This issue has already been pointed out in discussions of relational contracts. Indeed, it seems that relational contract theory also places a high reliance upon judicial capacity to evaluate and interpret norms,[162] despite their possible lack of expertise in particular commercial circumstances.[163] The difficulty is also exemplified by Fraser J's statement that, while the list of criteria was useful, it was not necessarily exhaustive.

These criticisms were tested when the High Court had to decide on relational contracts one year later. The test laid out in *Bates* was referred to and applied in *Cathay Pacific*,[164] but the court determined that a long-term aircraft engine agreement was not a relational contract because the spirit and objective of the venture had been clearly expressed in the contract and the parties did not rely on trust and confidence. The relationship was nothing more than a good working relationship, and the parties did not trust each other 'beyond what would normally expect in any commercial relationship'.[165] While the flexibility and adaptability of the notion of the relational contract have been brought forward against its development as a legal concept, the UK High Court showed that the test laid out in *Bates* could be articulated and applied to the facts before the court. While there is no one-size-fits-all approach, as Justice Fraser, throughout his judgment, refers to not only the terms of the contract, but also the context and

158 Joint Press Statement – Resolution to the Group Litigation Proceedings (Bates v Post Office Limited) (Post Office, 11 December 2019).

159 *Bates v Post Office (No 6)* [2019] EWHC 3408 (QB).

160 Ibid [936].

161 Ibid [693].

162 Ibid 153.

163 Macaulay (n 143 above) 66, 68, referring to Robert Scott on commercial law, common law and code methodologies, cited in Jody Kraus and Steven Walt, 'The jurisprudential foundations of corporate and commercial law' (2000) 94 Northwestern University Law Review 847. See also Lisa Bernstein, 'The questionable empirical basis of article 2's incorporation strategy: a preliminary study' (1999) 66 University of Chicago Law Review 710, 730.

164 *Cathay Pacific Airways Ltd v Lufthansa Technik AG* [2020] EWHC 1789 (Ch) [206], [225]–[236].

165 Ibid [232].

the relation of the parties seeing them as the key in determining the issues before the court. This flexible approach is key as context will change depending on the parties, the agreement and the context of the contract. And, according to Justice Fraser, this goes hand in hand with the duty to act in good faith. In *Taqa Bratani Ltd v Rockrose*, Pelling J did not refer to *Bates* and only quoted *Yam Seng* before stating that the contract was relational without attempting to define it further.[166] Pelling J also questioned whether such a designation necessarily entailed that a duty of good faith is implied. In *Essex County Council v UBB Waste (Essex) Limited*, Pepperall J used *Yam Seng* as well as the *indicia* laid out in *Bates* to determine that the contract in dispute was indeed relational by emphasising the long-term relationship, the close collaborative working relationship, the trust and confidence of the parties, the high degree of communication and cooperation, the significant investment by both parties and the exclusive nature of the agreement.[167] Consequently, a duty of good faith was implied.[168] A breach of the duty was examined objectively according to what would be considered 'commercially unacceptable by reasonable and honest people'.[169] In *Skeikh Tahnoon v Kent*, Lord Justice Leggatt further confirmed the relevance of the relational contract as a bridge between fiduciary relationships and discrete commercial transactions,[170] thereby further supporting the idea that contracts exist on a spectrum. In this instance, and relying on his judgment in *Yam Seng*, Lord Justice Leggatt found that an implied duty of good faith to a joint venture agreement recognised as relational had been breached.[171]

There has been a slow and incremental recognition of relational contracts as a category of contract law and of the consequences of implying a duty of good faith. Yet the contours of both this category of contracts and the implied term of good faith (if it exists at all) remain uncertain.[172] While decisions are still trying to articulate these concepts, the discussion in each of these decisions reflects norms of trust and solidarity, further illustrating McNeil's norms of contracting. While the above discussion has focused on case law and the UK, arguably a similar development is occurring in contract law in Australia but through targeted regulation instead of judicial decisions, which for

166 *Taqa Bratani Ltd v Rockrose* UKCS8 LLC [2020] EWHC 58 (Comm) [56].
167 *Essex County Council v UBB Waste (Essex) Limited* [2020] EWCH HC 1581 (TCC) [112]
168 Ibid [113].
169 Ibid [116]. The duty was not breached in this case.
170 *Sheikh Tahnoon Bin Saeed Bin Shakhboot Al Nehayan v Ioannis Kent* [2018] EWHC 333 (Comm) [167]
171 Ibid [176].
172 *UTB LLC v Sheffield United Limited* [2019] EWHC 2322 (Ch) [202].

now reflect a very conservative and classical approach to contracting in Australia.[173]

ACKNOWLEDGING THE IMPLICIT DIMENSIONS OF THE CONTRACT THROUGH INDUSTRY-SPECIFIC STATUTORY REGULATION IN AUSTRALIA

A slow recognition of good faith and relational contract theory seems to be taking place when regulating and adjudicating some long-term contracts. Some commercial contracts are deeply party-centric, and the relational contract approach acknowledges the relevance of this character, at least in some transactions. The main difference between the approach in Australia and that of the UK is the source of the renewed interest in relational contract theory. Yet, the commonality remains the incremental, slow and careful approach taken. Finn classifies good faith, unconscionability and fiduciary duties as part of a 'three-tier hierarchy of protective responsibility'.[174] Acknowledging and respecting the interests of the other party are common characteristics of this hierarchy,[175] albeit the fiduciary principle requires deference to the other party's interests. But there is a spectrum from the discrete transaction to the fiduciary principle. Relational contracts are within that spectrum but do not require parties to sacrifice their self-interest. Finding the balance between self-interest and consideration of the other party is arguably the most challenging part of articulating relational contracts and good faith. Over the last 10 years, Australia has articulated this dynamic tension through targeted legislation.

Industry-specific regulation such as codes of conduct and standards have helped answer industry-specific questions and legal issues. They consider the context and special characteristics of the relationship between the parties and its implicit dimensions. This is especially relevant in some ongoing, or long-term, business relations. Franchise agreements are long-term contracts whose relational characteristics initially led to a movement to imply a duty of good faith as a matter of

173 *Workpac Pty Ltd V Rossato & Ors* [2021] HCA 23; *Construction, Forestry, Maritime, Mining and Energy Union v Personnel Contracting Pty Ltd* [2022] HCA 1; *Zg Operations & Anor V Jamsek & Ors* [2022] HCA 2.

174 Finn (n 18 above) 142.

175 Ibid 136.

law.[176] The Franchising Code of Conduct is one of seven mandatory industry codes prescribed under the Competition and Consumer Act 2010 (Cth), section 51AE. Recognising 'the inherent and necessary imbalance of power in franchise agreements in favour of the franchisor, where abuse of this power can lead to opportunistic practice, a statutory duty to act in good faith'[177] would 'promote business integrity and ethics'.[178] Since its enactment in 2014, the Franchising Code of Conduct has included a duty on parties to act in good faith.[179] Since then, more codes of conduct have been regulated, and they also include a duty of good faith. Discussions on the recognition of a duty to act in good faith led to the reform of the mandatory Horticulture Code of Conduct[180] and the introduction of a new mandatory Dairy Code of Conduct.[181] The explicit purpose of the voluntary Food and Grocery Code of Conduct is 'to promote and support good faith in commercial dealings between retailers, wholesalers and suppliers'.[182] It imposes a duty to act in good faith at all times on the retailer and the supplier.[183] These examples further demonstrate what could be characterised as the particular need to regulate some long-term contracts, where the relationship of the parties is paramount. Each of these new industry codes also promotes good faith as an explicit enforceable obligation. These codes place the emphasis on notions of business integrity, ethics and the relationship between the parties. Good faith forms an integral part of this recognition. Each of these commercial contexts also have the potential to be considered relational, although the imbalance of power that is often found in these transactions is not in itself a criterion for a relational contract.

---

176 *Garry Rogers Motors (Aust) P/L v Subaru (Aust) P/L* (1999) 21 ATPR 41-703; *Burger King Corp v Hungry Jack's Pty Ltd* [2001] NSWCA 187; *Far Horizons Pty Ltd v McDonalds Australia Ltd* [2000] VSC 310; *AMC Commercial Cleaning v Coade* [2011] NSWSC 932; Bill Dixon, 'What is the content of the common law obligation of good faith in commercial franchises?' (2005) 33(3) Australian Business Law Review 207, 222.

177 Parliamentary Joint Committee on Corporations and Financial Services, *Opportunity not Opportunism: Improving Conduct in Australian Franchising* (Senate Printing Unit 2008) 101.

178 Tony Piccolo MP, Economics and Finance Committee, Parliament of South Australia, quoted in ibid 107.

179 Competition and Consumer (Industry Codes – Franchising) Regulation 2014 (Cth), cl 6; *Australian Competition and Consumer Commission v Ultra Tune Australia Pty Ltd* [2019] FCA 12.

180 Competition and Consumer (Industry Codes – Horticulture) Regulations 2017 (Cth).

181 Competition and Consumer (Industry Codes – Dairy) Regulations 2019 (Cth).

182 Competition and Consumer (Industry Codes – Food and Grocery) Regulation 2015 (Cth), s 2.

183 Ibid s 28(3).

Voluntary industry standards have also been drafted in the construction industry and a recent review of building standards has proposed a new explicit obligation on the principal and contractor 'to act reasonably in a spirit of mutual trust and cooperation, and generally in good faith towards the other'.[184] Here again the focus is on the legitimate interests of the parties and the need to come together with the ultimate aim to complete the agreement through performance. Unfortunately, the reform of these building standards has been put on hold indefinitely because it did not represent fully stakeholders' interests.[185] In late 2020, the adversarial nature of contracting in construction contracts was advanced as one of the reasons for the state of the building industry and the need for reform to embrace cooperation and collaboration.[186] In March 2021, the NEC 4 suite of contracts, well known in the UK, was introduced in the Australian construction industry. One of its core clauses contains a duty to act with mutual trust and cooperation, which has been associated with good faith.[187]

Applying *Yam Seng* to each of the transactions and relationships discussed in this section, they could arguably be said to include a

> high degree of communication, cooperation and predictable performance based on mutual trust and confidence and involve expectations of loyalty which are not legislated for in the express terms of the contract but are implicit in the parties' understanding and necessary to give business efficacy to the arrangements.[188]

The elements laid out in *Bates* also point towards these contracts being relational ones.[189] They are long-term relationships. The parties intend that their respective roles will be performed with integrity and with fidelity to their bargain. The parties are likely to be committed to collaborating with one another in the performance of the contract. There are implicit dimensions to their venture that may not be expressed in a written agreement. Each party is likely to repose trust and confidence in the other, without subordinating their

184 Building Standard AS11000:2015 cl 2.1. See also Alexander Di Stefano, 'Good faith in the AS11000: has the eagle landed?' (2016) 33 Building and Construction Law Journal 13; Joseph Biagio Xuereb, 'Is it time for an express term of good faith in Australian construction contracts' (2021) 36 Building and Construction Law Journal 229.

185 Standards Australia, Press Release (4 April 2017).

186 David Mosey, C*ollaborative Construction Procurement and Improved Value* (Wiley Blackwell 2019); Jennifer Clarkson, 'Covid-19 and construction law: comparing the UK and Australian response' (2021) International Construction Law Review 5.

187 *Mid Essex Hospital Services NHS Trust v Compass Group UK & Ireland Ltd (trading as Medirest)* [2013] EWCA Civ 200.

188 *Yam Seng PTE v International Trade Corp Ltd* [2013] EWHC 111 (QB) [142].

189 *Bates v Post Office Ltd (No 3)* [2019] EWHC 606 (QB) [725].

self-interest to the interests of the other party. These examples are likely to involve a high degree of communication, cooperation and predictable performance, based on mutual trust and confidence and expectations of loyalty. In most of these examples, there will be a degree of significant investment by one party and possibly exclusivity. Finally, not only is good faith not excluded, but through these codes of conduct it is expressly mandated. These examples demonstrate an incremental approach that is challenging classical contract theory in some well-defined situations. However, this piecemeal approach does not in itself justify changing Australian contract law principles.

CONCLUSION

With a renewed interest in the concept of the relational contract and the call for further empirical research,[190] Macneil's concept of the relational contract does have implications not only for contractual practice and management, but also for contract law itself. The parties to some long-term agreements and the written terms of the contract are surrounded by other forces that guide their relationship and their behaviour when performing their obligations. Taking the context and these forces into account leads to the recognition of the implicit dimensions of a contract. In spite of the classical hegemony of party autonomy in regulating their contractual terms, the need for fairness and justice in contract law has crossed over boundaries. The liberalist view, predominant in contract law, led to the classical theory of contract law and is exemplified by notions such as business efficacy, efficiency, freedom to contract and autonomy of the parties. Yet, morals have crept into the application of the law to soften its contours.

While the classic English contract law might have been the epitome of liberalism, current judicial developments demonstrate a change in the rationale for regulating contract law. A new species of contract is making an appearance, namely the relational contract. This notion, that was once relegated to academic debate, is now argued before the courts, and judges are actively using it as a lens to determine whether parties have breached their contractual obligations. The rationale for this movement can be found in the idea, summarised by Shand, that '[a] man must stick to his bargain, for otherwise social relations would not be possible'.[191] Parties who join in a contract must stick to the promises they have made. They have a duty to be faithful to the bargain.[192] It is reasonable for one party to the contract to expect

190 Cimino (n 80 above).

191 John Shand, 'Unblinkering the unruly horse: public policy in the law of contract' (1972) 30(1) Cambridge Law Journal 144, 147.

192 Finn (n 18 above) 148.

that the other party who entered the contract is willing to perform its obligation. Reasonable expectations go further in that they provide certain limits to the exercise of discretionary powers within the contract. Furthermore, a duty to act in good faith seems to be implied into some long-term transactions. The recent English case of *Bates* could have broad implications for contract law in that jurisdiction but also beyond. It does represent one of the stepping-stones to a doctrinal impact of the notion in contract law. Albeit in different ways, both the UK and Australia have had to regulate aspects of contract law to promote fairness in contractual dealings. While the approach in Australia has been more piecemeal and statutory, a bird's-eye view of the codes of conduct shows that the impetus to develop regulation that specifically targets some long-term contracts is growing.

What does this mean for the legal professional of today? Commercial contracts must be drafted more carefully, and lawyers must be aware, and possibly wary, of the consequence of a contract being considered relational by the courts: namely the impact on the interpretation and the difficulty of relying exclusively on the terms of the agreement. The context of the agreement and the relationship between the parties will matter, although to what extent is still an open question.

Northern Ireland
Legal Quarterly
Autumn Vol. 74.3 (2023) 559–587
Articles DOI: 10.53386/nilq.v74i3.1120

The impact of proposed intimate image abuse offences on domestic violence and abuse†

Charlotte Bishop

University of Reading
Correspondence email: c.bishop@reading.ac.uk

ABSTRACT

This article explores how proposed reforms to the law on intimate image abuse could address situations where intimate images are shared, or threats to share are made, in a relationship where there is domestic violence and abuse (DVA). In exploring the purposes and motivations behind the use of non-consensual intimate images in this context, the harmful impact is demonstrated to be the denial of autonomy and personhood that 'entraps' the victim in the relationship. It is essential that this harm, and the underlying motivations of those who use intimate image abuse for this purpose, is made visible under any relevant legislation to ensure that the criminal law effectively condemns and remedies conduct of this kind. It is for this reason that the article concludes that the Law Commission, in its 2021 consultation, was right to consider introducing an offence of 'intentionally taking or sharing an intimate image without consent with the intent to control or coerce the person depicted'.[1] It is further suggested that this fault element may better reflect the culpability of those who engage in *threats* to share intimate images and should be introduced not just where images are taken and shared, but also where threats to share such images are made.

Keywords: intimate image abuse; gender-based violence; domestic violence; domestic abuse; coercive control; law reform.

INTRODUCTION

In November 2018 the Law Commission noted a compelling need for review and reform of the law on the non-consensual taking and sharing of intimate images.[2] A project then commenced in June 2019 to review the current offences in this area, identify gaps in the scope of the protection currently offered, and make recommendations to ensure the criminal law provides consistent and effective protection against

† First published in NILQ 73.AD2 125–153 6 October 2022.

1 Law Commission, *Intimate Image Abuse: A Consultation Paper* (Law Com No 253, 2021) at para 10.93.

2 Law Commission, *Abusive and Offensive Online Communications: A Scoping Report* (Law Com No 381, 2018).

the creation and sharing of intimate images without consent.[3] As part of this review, a consultation seeking feedback on provisional reform proposals was published in February 2021.[4] This article develops responses provided by the author to specific aspects of the proposed reforms as they would relate to intimate image abuse[5] that occurs within the context of domestic violence and abuse (DVA).[6] While this is not the only area in which urgent reform is needed, the prevalence of non-consensual taking, making and sharing of intimate images in this context makes this an important area of focus.[7] The harm that results from intimate image abuse is serious, far-reaching, and long-lasting, extending from distress and humiliation to psychological trauma, anxiety and depression, suicidal ideation, loss of employment, and, in extreme cases, suicide.[8] In the context of DVA, sharing and/or

3 Law Commission (n 1 above) para 1.9 and 'Taking, making and sharing intimate images without consent'.

4 Law Commission (n 1 above).

5 This is the term adopted by the Law Commission in its current consultation on 'Taking, making and sharing intimate images without consent' (n 3 above). Terminology is discussed below.

6 There is a discussion on terminology below with reference to ongoing debates with regards the most effective term to use in referring to violence and abuse against intimate partners.

7 While it is estimated that globally between 8% and 13% of individuals have experienced intimate image abuse (A Eaton et al, *2017 Nationwide Online Survey of Nonconsensual Porn Victimization and Perpetration* (Cyber Civil Rights Initiative June 2017) ; '8 percent of Brits are victims of "revenge porn"' (*Open Access Government* 21 March 2019), there is a strong relationship between intimate image abuse, domestic abuse and coercive control (N Henry et al, *Image-Based Sexual Abuse: A Study on the Causes and Consequences of Non-Consensual Nude or Sexual Imagery* (Routledge 2021). Threats to share intimate images are particularly common in this context. Research by Refuge indicates that 1 in 7 young women have experienced threats, with 72% stating that these were from a current or former partner (*The Naked Threat Report* (Refuge July 2020)). The Revenge Porn Helpline reports that 1 in 4 of the calls it receives relate to threats (Law Commission (n 1 above) para 3.76).

8 C McGlynn et al, '"It's torture for the soul": the harms of image-based sexual abuse' (2020) 30 Social and Legal Studies 541; C McGlynn et al, *Shattering Lives and Myths: A Report on Image-Based Sexual Abuse* (Durham University and the University of Kent July 2019); C McGlynn and E Rackley, 'Image-based sexual abuse' (2017) 37 Oxford Journal of Legal Studies 534; C McGlynn et al, 'Beyond "revenge porn": the continuum of image-based sexual abuse' (2017) 25 Feminist Legal Studies 25; D K Citron and M A Franks, 'Criminalising revenge porn' (2014) 49 Wake Forest Law Review 345; A Powell and N Henry, *Sexual Violence in a Digital Age* (Palgrave Macmillan 2017) ch 5; T Crofts and T Kirchengast, 'A ladder approach to criminalising revenge pornography' (2019) 83 Journal of Criminal Law 87; S Bates, 'Revenge porn and mental health: a qualitative analysis of the mental health effects of revenge porn on female survivors' (2017) 12 Feminist Criminology 22.

threatening to share intimate images have become a central component of the tools used by perpetrators to maintain power and control and prevent victims from leaving the relationship.[9] The harm that results from this extends beyond that which arises from the act of sharing or from fear that the threat to share the image will be realised.[10] Here intimate images are used to disempower the victim, thus depriving them of autonomy and personhood and leaving them isolated and dependent on the very person who is abusing them.[11] These harms make it vital that this aspect of intimate image abuse is effectively captured in any new legal framework that is introduced.

This article begins by discussing the societal and cultural context in which intimate image abuse must be understood and the best terminology to be used when discussing its role within abusive relationships. It then outlines the ways in which sharing and threatening to share intimate images have become tools of coercion and control within intimate relationships and articulates the importance of understanding the motivation behind this behaviour, as well as the harmful loss of autonomy it results in, when perpetrator culpability is determined. Following this, the limitations of the existing criminal law relating to intimate images are outlined, underlining the need for legal reforms in this area to provide more consistent protection and ensure victim experiences and perpetrator motivations are adequately reflected in legal definitions and considered appropriately at sentencing. The article then discusses the offences proposed in the consultation and makes recommendations about how these issues of motive and impact could usefully be captured in assessing culpability and determining the fault elements.

---

9 N Henry and A Powell, 'Beyond the "sext": technology-facilitated sexual violence and harassment against adult women' (2015) 48 Australian and New Zealand Journal of Criminology 113; A Eaton et al, 'Nonconsensual porn as a form of intimate partner violence: using the power and control wheel to understand nonconsensual porn perpetration in intimate relationships' (2021) 22 Trauma, Violence and Abuse 1140; Citron and Franks (n 8 above); M Dragiewicz et al, 'Technology facilitated coercive control: domestic violence and the competing roles of digital media platforms' (2018) 18 Feminist Media Studies 609; C Dardis and E Richards, 'Nonconsensual distribution of sexually explicit images within a context of coercive control: frequency, characteristics, and associations with other forms of victimization' (2022) Violence Against Women 1; Henry et al (n 7 above).

10 D Cuomo and N Dolci, 'New tools, old abuse: technology-enabled coercive control (TECC)' (2021) 126 Geoforum 224, 230.

11 E Stark, *Coercive Control: How Men Entrap Women in Personal Life* (Oxford University Press 2007); V Tadros, 'The distinctiveness of domestic abuse: a freedom based account' (2004) 65 Louisiana Law Review 989.

CONTEXT AND TERMINOLOGY

A rapidly growing body of research demonstrates that the vast majority of images being shared and traded online are of women and that intimate image abuse is committed mainly by men.[12] Most frequently this is a current or former romantic partner and the abuse often co-occurs with offline forms of male-to-female assault.[13] However, because both DVA and intimate image abuse are part of the broader constellation of gender-based violence[14] it is not just that women and girls are statistically more likely to be victims, but that both phenomena are 'qualitatively gendered'. As McGlynn and Rackley emphasise, the harms of intimate image abuse are gendered due to the 'sexualised and misogynistic form and manner in which they are manifested' and the sexual double standards that prevail at a societal and cultural level.[15] They draw attention to societal gender disparities that enable and facilitate the production and prevalence of intimate image abuse.[16] The same can be said of the prevalence and nature of DVA. When understood as a programme of coercive control intended to disempower the victim, rather than as 'incidents' of physical violence and other types of abuse, DVA can be seen as a gendered social phenomenon that results from structural inequality and is delivered via the exploitation of gender norms. The behaviours characteristic of coercive control focus on the micro-regulation of many of the everyday activities and roles already typically associated with women in their roles as homemakers, parents

12 C Uhl et al, 'An examination of nonconsensual pornography websites' (2018) 26 Feminism and Psychology 50; J Halliday, 'Revenge porn: 175 cases reported to police in six months' *The Guardian* (Manchester, 11 October 2015). Figures from the UK's Revenge Porn Helpline show that 75% of 1800 calls over six months were from women: Government Equalities Office, 'Hundreds of victim-survivors of revenge porn seek support from helpline' (Press Release 20 August 2015).

13 N Henry and A Powell, 'Embodied harms: gender, shame and technology-facilitated sexual violence' (2015) 21 Violence Against Women 758, 760; W DeKeseredy, 'Image-based sexual abuse: social and legal implications' (2021) 8 Current Addiction Reports 330, 330; Y Ruvalcaba and A Eaton, 'Nonconsensual pornography among US adults: a sexual scripts framework on victimization, perpetration, and health correlates for women and men' (2020) 10 Psychology of Violence 68; M Hall and J Hearn, 'Revenge pornography and manhood acts: a discourse analysis of perpetrators' accounts' (2017) 28 Journal of Gender Studies 158; N Henry and A Flynn, 'Image-based sexual abuse: online distribution channels and illicit communities of support' (2019) 25 Violence Against Women 1932.

14 Gender-based violence is defined by the UN as 'harmful acts directed at an individual based on their gender and rooted in gender inequality, the abuse of power and harmful norms': United Nations, 'Gender-based violence' (UN Refugee Agency).

15 McGlynn and Rackley (n 8 above) 544.

16 Ibid.

and sexual partners.[17] This means women are held accountable for their performance of femininity at the same time as masculinity is reinforced through the normalisation of male power and control. Stark explains the rise of coercive control in recent decades as resulting from women's formal equality gains in the public sphere because the dismantling of uncontested male power at a societal level has left men needing to bolster masculine identity by developing more overt ways of controlling the *individual* women in their personal lives.[18] However, the rise in the public sharing of private images of women, alongside accompanying text that objectifies, shames and humiliates them, can be positioned as a contemporary device that is enabling the re-normalisation of male power and control over women in the public sphere. As Nussbaum observes, 'the online objectification of women' can be seen as some men's attempts to 'restor[e] the patriarchal world before the advent of sex equality, the world in which women were just tools of male purposes'.[19] Collectively, women are 'kept in their place' through the pervasive normalisation and eroticisation of male dominance and female subordination that occurs through the non-consensual sharing of explicit and intimate images in online public spaces. In this way, distributing or threatening to distribute intimate images is a highly effective tactic for disciplining victims and keeping them in abusive relationships, since they know that ongoing systemic sexism will position them as responsible for the impact of the abuse.[20]

While practical suggestions for addressing the wider collective harm of intimate image abuse fall beyond the scope of this article, it is essential to acknowledge the public nature of the wrong to ensure the legal response is premised upon an assessment of the seriousness of wrongdoing that takes account of the harm to individual and *collective* interests.[21] As highlighted by Von Hirsch and Jareborg, collective

17 Stark (n 11 above) 129–130. Also see K Anderson, 'Gendering coercive control' (2009) 15 Violence Against Women 1444.

18 Stark (n 11 above).

19 M Nussbaum, 'Objectification and ressentiment', unpublished paper on file with Danielle Keats Citron and quoted in D K Citron, 'Law's expressive value in combating cyber harassment' (2009–2010) 108 Michigan Law Review 373, 389.

20 Cuomo and Dolci (n 10 above) 230. For discussions of victim-blaming in this context, see McGlynn and Rackley (n 8 above); S Bothamley and R Tully, 'Understanding revenge pornography: public perceptions of revenge pornography and victim blaming' (2018) 10 Journal of Aggression, Conflict and Peace Research 1; Crofts and Kirchengast (n 8 above); Hall and Hearn (n 13 above); E Rackley et al, 'Seeking justice and redress for victim survivors of image based sexual abuse' (2021) 29 Feminist Legal Studies 293; A Powell et al, 'Image-based sexual abuse: the extent, nature, and predictors of perpetration in a community sample of Australian residents' (2019) 92 Computers in Human Behavior 393; Henry et al (n 7 above); Eaton et al (n 9 above).

21 Crofts and Kirchengast (n 8 above) 90.

interests are relevant when assessing harm.[22] Ashworth views crimes as public wrongs, even where they are an attack on an individual, as with intimate image abuse, because they are 'wrongs that are shared by other members of the community with which the victim is identified and by which her or his identity is partly constituted'.[23] Therefore, when considering how to ensure legal reforms are as effective as possible, both intimate image abuse and DVA must be situated within their wider social and cultural context to enable the impact and the harm caused to victims, and the culpability of those who engage in it, to be appropriately captured in legislation and wider government and criminal justice policy.

Intimate image abuse

Different terms have been used to describe the non-consensual sharing of sexual and/or intimate images, some of which have proved problematic and controversial. Commonly used by the media and within official guidance,[24] the term 'revenge porn' is widely criticised as oversimplifying and misrepresenting perpetrator motivations and victim experiences.[25] It reduces the severe harms that result from the non-consensual publishing of intimate images to a 'simple "scorned ex-boyfriend" narrative' and suggests that perpetrators are motivated only by personal vengeance.[26] This implies that victims must have done something to cause the perpetrator to seek 'revenge', thus perpetuating victim-blaming and shifting focus away from the motivations and behaviours of the perpetrator and on to the content of the images and the actions of the victim.[27] Popular use of such an inadequate term has shaped the development of new laws around the world and influenced police responses to victims.[28] For example, in England and Wales prior use of the term 'revenge porn' in legislative debates narrowed the scope of discussions and limited the applicability of the 'disclosure'

22 A Von Hirsch and N Jareborg, 'Gauging criminal harm: a living-standard analysis' (1991) 11 Oxford Journal of Legal Studies 1, 33

23 A Ashworth, 'Is the criminal law a lost cause?' (2000) 116 Law Quarterly Review 225, 243

24 See Ministry of Justice promotional materials, 'Revenge porn: be aware b4 you share'; Ministry of Justice, 'Revenge porn: the facts' (2021); Crown Prosecution Service, 'Social media – guidelines on prosecuting cases involving communications sent via social media' (2018); College of Policing, 'Revenge pornography' (2020).

25 Henry et al (n 7 above).

26 S Maddocks, 'From non-consensual pornography to image-based sexual abuse: charting the course of a problem with many names' (2018) 33 Australian Feminist Studies 345, 347.

27 McGlynn and Rackley (n 8 above) 536.

28 Powell and Henry, *Sexual Violence* (n 8 above) 203.

offence[29] to those seeking to cause distress.[30] The examples provided in a recent study by Henry et al show a diverse range of motivations that often overlap and intersect, with power and control commonly the overarching theme.[31]

Franks coined the term 'non-consensual pornography' (NCP) in an attempt to foreground the underlying disregard for women's consent.[32] However, as with 'revenge porn', categorising the sharing of intimate images as a 'subgenre of commercially produced online pornography' is problematic. It implies an element of choice and autonomy on the part of the victim and overlooks the fact the images were not created for public consumption, may not even have been *taken* with the consent of the victim, and are often not shared or accessed with the primary aim of sexual gratification.[33] More recently, McGlynn and Rackley proposed the term image-based sexual abuse (IBSA), defined as 'the non-consensual creation and/or distribution of private, sexual images'.[34] Drawing on Kelly's notion of a 'continuum of sexual violence',[35] IBSA is conceptualised as one harm situated along a continuum of sexual abuses – from catcalling to rape – driven by the same societal disregard for women's consent.[36] NCP and IBSA cover an almost identical range of harms, their difference is in emphasis: NCP refers to the product (pornography), while IBSA describes the conduct and its impact on victims (abuse).[37] While there are clear advantages to the term IBSA, particularly over terms such as revenge porn or NCP, the Law Commission consultation chose to use the term 'intimate image abuse' on the basis that this is an inclusive term, encompassing both the nature of the images under consideration and the range of harmful behaviours demonstrated by perpetrators.[38] It is agreed that this is a favourable term since 'intimate' denotes the nature of the images and retention of the word 'abuse' ensures, as with IBSA, that the impact on victims is centred. At the same time, as noted in the consultation, it acknowledges and reflects the lack of consensus

29 S 33 Criminal Justice and Courts Act 2015.

30 McGlynn and Rackley (n 8 above) 553.

31 Henry et al (n 7 above) 85.

32 M A Franks, 'An-aesthetic theory: Adorno, sexuality, and memory' in R Heberle (ed), *Feminist Interpretations of Theodor Adorno* (Pennsylvania State University Press 2016).

33 Maddocks (n 26 above) 349 suggests that the explanatory power of the term 'pornography' should not be disregarded since porn sites have become the main repositories for non-consensually distributed sexual content.

34 McGlynn et al, 'Beyond "revenge porn"' (n 8 above).

35 L Kelly, *Surviving Sexual Violence* (1988 Polity Press).

36 This is discussed further in McGlynn et al, 'Beyond "revenge porn"' (n 8 above).

37 Maddocks (n 26 above) 350.

38 Law Commission (n 1 above) at para 1.13.

on whether all examples of this behaviour should be identified and punished as sexual offending.[39]

Victim/survivor

The term 'survivor' is now often preferred over the label 'victim,' in recognition of the agency and coping capacities of women who have experienced gender-based violence.[40] However, in the present context it may reduce the persuasiveness of the arguments to use the term 'survivor', with the agency and free will this implies, since the article is arguing for attention to be paid to the ways in which intimate images are being used to deprive women of autonomy and 'entrap' them in coercively controlling relationships. 'Staying' in the abusive relationship must also not be constructed as being driven by 'choice' in any meaningful sense and it must be remembered that coercive control is often ongoing even where the relationship has officially ended.[41] It is also consistent in an article on reforms to the criminal law to refer to the person against whom an offence has been committed as a 'victim,' since this is the term used by criminal justice system agencies (as well as the 'complainant', prior to conviction). Therefore, the term 'victim' is retained, while at the same time it is recognised that it is neither accurate nor desirable to present women as passive victims given the myriad strategies and tactics of resistance that they engage in on a daily basis while navigating male violence and control.[42]

Domestic violence and abuse

Debates on how to refer to and define violence and abuse against intimate partners are well rehearsed and long running.[43] A range of different terms are used and are often reflective of the particular theoretical and methodological approach used, the type of violence and abuse being studied, and the contexts that the researcher wishes to foreground.[44] In the United Kingdom, domestic violence was the most commonly used term until relatively recently when there has been a

39 Law Commission (n 3 above) at para 1.15.

40 Kelly (n 35 above).

41 C Wiener, 'From social construct to legal innovation: the offence of controlling or coercive behaviour in England and Wales' in M McMahon and P McGorrey (eds), *Criminalising Coercive Control* (Springer 2020); D Tuerkheimer, 'Breakups' (2013) 25 Yale Journal of Law and Feminism 51; H Douglas, 'Legal systems abuse and coercive control' (2018) 18 Criminology and Criminal Justice 84.

42 Stark (n 11 above); M Dutton and L Goodman, 'Coercion in intimate partner violence: toward a new conceptualization' (2005) 52 Sex Roles 743; C Wiener, 'What is "invisible in plain sight": policing coercive control' (2017) 56 (4) Howard Journal of Crime and Justice 500.

43 Dragiewicz et al (n 9 above) 610.

44 Ibid.

shift towards 'domestic abuse' in law and official policy.[45] Aldridge notes the problematic and unusual nature of this move and suggests this could be a deliberate move to underplay DVA as a gendered issue, with the removal of 'violence' as a key rubric suggesting a 'watering-down' or obfuscation of the serious and gendered nature of DVA.[46] DVA is therefore the term used in this article to facilitate focus on a broader range of harms, but at the same time it is acknowledged that no term is perfect; each has its limitations and advantages.[47]

As will be explored more fully below, DVA is best understood as a 'liberty crime' in which abusers entrap victims, undermine their social support, subvert their autonomy and deprive them of equality.[48] The inclusion of 'coercive control' as a distinct 'type' of DVA is therefore regarded to be erroneous.[49] In recent years, researchers have emphasised the role of digital technologies in facilitating this pattern of abuse.[50] Technology and digital media offer a variety of everyday options for effectively controlling partners in often similar ways to traditional forms of abuse, such as stalking and surveillance, but the accessibility and immediacy of mobile, digital and social media may result in abuse perpetration with greater ease, using new methods and channels.[51] Therefore, terms such as 'technology-enabled' and 'technology-facilitated' coercive control (TECC and TFCC) usefully draw attention to the range of abusive behaviours that are carried out using digital technology, of which intimate image abuse is clearly only a part.[52] However, for the purposes of this article it is not considered useful to break behaviour down into types of coercive control depending on whether it takes place in an online or offline space. Instead, it is most

45 For example, in the recently passed Domestic Abuse Act 2021 and accompanying statutory guidance.

46 J Aldridge, '"Not an either/or situation": the minimization of violence against women in United Kingdom "domestic abuse" policy' (2020) Violence Against Women 2.

47 Dragiewicz et al (n 9 above) 610.

48 Stark (n 11 above); Tadros (n 11 above).

49 This approach is found in the definition of 'domestic abuse' contained in s 1 of the Domestic Abuse Act 2021.

50 Cuomo and Dolci (n 10 above); Dragiewicz et al (n 9 above).

51 D Woodlock, 'The abuse of technology in domestic violence and stalking' (2017) 23 Violence Against Women 584–602; Cuomo and Dolci (n 10 above) 224.

52 Researchers note that digital technologies form part of a 'constellation of tactics' that perpetrators of domestic abuse employ alongside other more widely recognised forms of physical and psychological abuse that take place in face-to-face encounters (H Douglas et al, 'Technology-facilitated domestic and family violence: women's experiences' (2019) 59 British Journal of Criminology 551; L Reed et al, 'Snooping and sexting: digital media as a context for dating aggression and abuse among college students' (2016) 22 Violence Against Women 1556–1576).

useful to use the term DVA – defined as a programme of coercively controlling behaviours aimed at disempowering the victim – and to then consider perpetrator motivations and intentions when engaging in this behaviour, for the purposes of assessing offender culpability.

INTIMATE IMAGE ABUSE AS A TOOL OF COERCIVE CONTROL

A number of scholars have already described and examined relationship-based intimate image abuse as a form of DVA, and it is clear that taking, making, sharing and threatening to share sexual and intimate images are now integral parts of the constellation of behaviours carried out by perpetrators in abusive relationships.[53] In assessing how intimate images are used in abusive relationships, and ensuring the harm to the victim and the culpability of the defendant are captured in any legal reforms in this area, it is necessary to understand how DVA itself typically manifests and how intimate images are then used in this context.

In recent years, the conceptual framework of coercive control, developed by Stark, has been used by feminist scholars in an attempt to shift the collective understanding of DVA away from a narrow focus on decontextualised acts of physical violence. By foregrounding patterns of behaviour and the constellation of abusive tactics used to entrap partners and limit their freedom and autonomy, DVA can be seen as a systematic process of coercive control intended to disempower the victim. In moving away from the dominant incident-based approach, the use of physical violence and other 'episodes' of abuse become understood as tools used to disempower the victim, rather than as articulations of the harm in and of themselves.[54] Although a set of discrete abusive incidents can typically be identified within an abusive relationship, DVA is not episodic; these incidents are connected by dynamics of power and control.[55] As Dutton has emphasised, 'to negate the impact of the time period between discrete episodes of serious violence … is to fail to recognize what some battered women experience as a continuing

53 Eaton et al (n 9 above); Henry and Powell, 'Beyond the "sext"' (n 9 above); Henry and Powell, 'Embodied harms' (n 13 above); Powell and Henry, *Sexual Violence* (n 8 above); Reed et al (n 52 above); Refuge (n 7 above); Citron and Franks (n 8 above); Cuomo and Dolci (n 10 above); Dardis and Richards (n 9 above); Dragiewicz et al (n 9 above).

54 Stark (n 11 above).

55 Tuerkheimer (n 41) 52.

"state of siege"'.[56] The victim never knows when the next episode will occur and lives in a permanent state of hypervigilance, as reflected by victims who describe DVA as an ongoing, 'everyday' reality in which much of their behaviour is 'micro-managed' by their abuser.[57] Once attention is drawn to the overall impact of the myriad controlling, violent and abusive behaviours carried out by the perpetrators, the ways in which victims are entrapped in the relationship become evident and the question 'why didn't she just leave?' becomes obsolete. Understandings of the harm, and the motivation of the perpetrator, are also transformed.

The concept of 'separation assault' describes a particular type of assault 'on a woman's body and volition that seeks to block her from leaving, retaliate for her departure, or forcibly end the separation'.[58] Separation assaults commonly occur when the victim makes an attempt to leave, or expresses the desire to end the relationship, meaning a great deal of 'separation' assaults take place *during* the relationship in an attempt to regain and maintain power and control over the victim. This concept shifts the focus on to what the perpetrator is doing that is limiting the victim's autonomy and makes visible the dynamics and behaviours that are keeping the victim entrapped. Studies draw attention to the routine use of 'credible threats' by the perpetrator to maintain power and control over the victim.[59] These may be threats of rape or physical violence, threats to take children away or report the victim to the authorities, or threats to humiliate the victim in public or in front of family, friends, or work colleagues. Importantly, they are used when the victim attempts to leave or to assert their autonomy, and therefore each threat functions as a means by which the perpetrator entraps the victim in the relationship and ensures compliance with the demands and rules they impose.[60] Each threat gains credibility, and therefore enables the perpetrator to maintain control, because the victim knows, based on past experience, that the perpetrator has the means and motivation to carry out the threat.

Given the ease with which images can now be taken and distributed, and the humiliation and harm that results, taking, sharing and threatening to share intimate images have become very effective tools of coercive control, often serving as 'credible threats' used by the perpetrator to entrap the victim and deprive them of autonomy.

56 M A Dutton, 'Understanding women's responses to domestic violence: a redefinition of battered woman syndrome' (1993) 21 Hofstra Law Review 1191, 1208.

57 Stark (n 11 above).

58 M Mahoney, 'Legal images of battered women: redefining the issue of separation' (1991) 90 Michigan Law Review 1, 6.

59 Dutton and Goodman (n 42 above); Wiener (n 42).

60 Dutton and Goodman (n 42 above).

Sometimes the images are initially shared or captured consensually, during positive periods of the relationship.[61] Other times images are obtained under coercion or are obtained covertly through the use of hidden cameras, secretly recording webcam communications, or using other forms of surveillance that the survivor was not aware of at the time.[62] The threat of dissemination can be used to intimidate victims, ensure compliance with demands, prevent them from leaving or reporting abuse, and may be retained for future use in case coercion or blackmail is needed.[63] Research by Refuge indicates that threats to share leave women feeling forced into telling perpetrators where they are, resuming or continuing the relationship, or allowing contact with children.[64] Abuse in a relationship is context-specific and built and maintained over time, such that, over time, fewer threats are needed to ensure compliance, and the carrying out of credible threats in the past means the victim has good reason to believe that the perpetrator will follow through with the threat in future, leading to greater vulnerability to threats or demands based around sexual images in the future.[65]

Previous research has theorised intimate image abuse as a form of DVA, based on feminist theories of power and control. Eaton and colleagues conceptualised the tactics of intimate image abuse (or NCP as they refer to it) by applying the power and control wheel[66] to illustrate the interconnected tactics used by abusers to assert dominance and control. In doing this, they evidence that NCP and DVA are perpetrated using similar tactics, and thus co-occur within an 'interlocking pattern of abuse'.[67] As emphasised by Dardis and Richards, this theory complements the framework of coercive control and credible threats outlined above; the power and control wheel specifies tactics used within patterns of coercive control that are established more strongly over time with repeated and varied forms of violence. However, in calling for NCP in relationships to be treated as a potential 'form' of partner violence, it is important not to reinforce understandings of DVA as consisting of different 'types' of abuse. For present purposes, it is therefore more useful, and more in line with theorising by Dragiewicz et al discussed above, to conceptualise the

61 Cuomo and Dolci (n 10 above).

62 Ibid 229; Henry et al (n 7 above) 81.

63 Eaton et al (n 9 above); Citron and Franks (n 8 above); Refuge (n 7 above). Law Commission (n 3 above) para 12.3.

64 Refuge (n 7 above).

65 Dutton and Goodman (n 42); Stark (n 11 above). Also see Dragiewicz et al (n 9 above); Dardis and Richards (n 9 above); Cuomo and Dolci (n 10 above).

66 The power and control wheel was developed in the 1980s (E Pence and M Paymar, *Education Groups for Men who Batter* (Springer 1993)) to characterise an interrelated and interlocking system of abusive and violent behaviours.

67 Eaton et al (n 9 above) 11.

various ways intimate image abuse manifests in abusive relationships as tactics of coercive control, rather than seeking to show that intimate image abuse is perpetrated via coercive control tactics.

When considering the role of intimate image abuse in DVA, it is not that intimate image abuse must now be seen as a 'type' of abusive behaviour, it is that the use of intimate images must now be recognised as part of a constellation of behaviours used to coercively control victims and deprive them of their autonomy. This focuses on the purpose intimate images are used for, and the overall pattern of abusive behaviour, rather than on specific aspects of it, therefore enabling articulation of harm and corresponding culpability. Referring to 'types' of abusive behaviour isolates and decontextualises the acts from the relational context in which they occur, thus obscuring the meaning, underlying motivation, and impact on the victim that the behaviour has. This leads to a rupture between women's experiences and the remediation offered by the criminal law; the law cannot fully condemn or remedy harm that it does not recognise.[68] The functioning of the system, and the justification for imposing punishment, is contingent on its proper recognition of harm and therefore, if the criminal law does not accurately capture the experience of the victim then its legitimacy is severely undermined.[69] Although a full critique of the offence of 'controlling or coercive behaviour' under section 76 of the Serious Crime Act 2015 (hereafter s 76 offence) is not possible here, legal scholarship has demonstrated the limitations of this offence because of its misconstruction of the nature and harm of coercive control.[70] The legal reforms currently under discussion present an opportunity to learn from this mistake by ensuring harm and corresponding culpability are appropriately reflected in the proposed offences.

CURRENT LAW

At the moment no criminal offence comprehensively covers the taking, making and sharing of intimate images without consent. There are three specific offences that may apply to some forms of these behaviours and a number of other offences that may also be used in this context. None are specific to situations of DVA, but all *could* be relevant in this context, depending on the circumstances. There is also no specific offence that criminalises all forms of *threats* to take, make and share intimate images without consent, but a number of the offences could

68 D Tuerkheimer, 'Recognising and remedying the harm of battering: a call to criminalize domestic violence' (2004) 94 Journal of Criminal Law and Criminology 970, 961.

69 Ibid 1015.

70 Wiener (n 41); Wiener (n 42).

be used to address threats of this kind. Each of these existing offences is dealt with in turn.

Section 33 Criminal Justice and Courts Act 2015 criminalises the disclosure of private sexual photographs or films without consent with intent to cause that individual distress (the 'disclosure' offence). However, it is not widely used and the attrition rate is high.[71] The specific intent requirement, whereby it must be proven that the defendant intended to cause distress to the victim, operates as a significant evidential barrier.[72] The offence is also categorised as a communications offence which limits the maximum sentence to 2 years on conviction in the Crown Court and 6 months in the magistrates' court, as well as denying victims the protections afforded them in sexual offences cases.[73] Where images are shared as a tool of coercive control, 'intention to cause distress' does not fully capture the motivations of the perpetrator or the resultant harm to the victim, which can be far more devastating than mere 'distress'.

Following a campaign by Refuge highlighting the prevalence of threats, the disclosure offence was amended in 2021 to include threats to disclose intimate sexual images.[74] When this amendment was proposed, the primary harm targeted was identified to be coercion and control in an abusive relationship.[75] However, in mirroring the existing offence, it suffers the same limitations and particularly, with the same specific fault element, does not adequately capture the motivation behind this behaviour when used as a tool of coercion and control in intimate relationships.

The voyeurism offences under the Sexual Offences Act 2003 (SOA 2003) can be applied to intimate image abuse, provided the

71 According to data provided to the Law Commission in October 2020, prosecutions have been falling since 2018 and no action was taken in 64% of reported offences, the main reasons given being a 'lack of evidence' and 'the victim withdrawing support' (Law Commission (n 1 above) para 3.8).

72 Ibid paras 3.46 and 3.63–3.66.

73 These measures include the granting of automatic anonymity under s 1 Sexual Offences (Amendment) Act 1992, special measures in court under ss 16–30 Youth Justice and Criminal Evidence Act 1999, and limitations on cross-examination are required following a line of cases concerning changes to the questioning practices as regards vulnerable witnesses (*Barker, E* [2011] EWCA Crim 3028; *W* [2010] EWCA Crim 1926; *Wills* [2011] EWCA Crim 1938). See also E Henderson, 'Taking control of cross-examination: judges, advocates and intermediaries discuss judicial management of the cross-examination of vulnerable people' (2016) Criminal Law Review 181.

74 S 69 Domestic Abuse Act 2021.

75 In introducing the proposed amendment, Baroness Morgan of Cotes stated: 'this is an issue about the exercise of control by one person—the abuser, the maker of the threats—over another. Too often, the threats are followed by physical abuse.' (Hansard, HL Deb, 8 February 2021, vol 810, col 145).

defendant acted with the purpose of obtaining sexual gratification.[76] The 'upskirting' offence may also be used, provided the *actus reus* is fulfilled.[77] This offence extends to situations where the perpetrator is acting with the purpose of humiliating, alarming or distressing the individual in the images, rather than being limited to sexual gratification. However, this specific purpose must be proven by the prosecution, which may cause similar difficulties to those outlined above with regards to the 'disclosure' offence. The offence also only applies to the taking of images, not where images are subsequently shared, and only applies to this particular type of behaviour so is narrow in scope and unlikely to be used in the context of DVA.

Some intimate image abuse, including threats to share images, may fall under the stalking and harassment offences contained in the Protection from Harassment Act 1997.[78] For both the stalking and harassment offences there must be a 'course of conduct', defined as conduct on at least two occasions.[79] This may be challenging for prosecutors, since in many cases there is likely only to be a small number of incidents; a threat to share an intimate image may only need to be made once to create the effect desired by the perpetrator. If a threat is not repeated, it may be difficult to substantiate a course of conduct and therefore apply the offence.[80] The legislation also does not apply where a relationship is ongoing, meaning it cannot be used where intimate image abuse, including threats, happens in this context.[81] Harassment is defined in section 7(1) to include 'alarming the person or causing the person distress' and has been further defined by Lord Sumption as a 'persistent and deliberate course of unreasonable and oppressive conduct, targeted at another person, which is calculated to and does cause that person alarm, fear or distress'.[82] Although intimate image abuse could often constitute conduct of this kind, the proof of harm

76 There are four voyeurism offences under s 67 SOA 2003, all of which are designed to criminalise the observing or recording of private acts of others without their consent.

77 The practice of 'upskirting' involves operating equipment under a person's clothing without their consent, with the intention of viewing their genitals or buttocks, with or without underwear. The offence was introduced under s 67A SOA 2003 by the Voyeurism (Offences) Act 2019.

78 The relevant offences are s 2 (harassment); s 2A (stalking); s 4 (putting people in fear of violence); s 4A (stalking involving fear of violence or serious alarm or distress).

79 S 7(3).

80 Law Commission (n 1 above) para 12.41.

81 *Curtis* [2010] EWCA Crim 123; [2010] 3 All ER 849; *Widdows* [2011] EWCA Crim 1500; [2011] Crim LR 959. See also V Bettinson and C Bishop, 'Is the creation of a discrete offence of coercive control necessary to combat domestic violence?' (2015) 66 Northern Ireland Legal Quarterly 179–197, 187–90.

82 *Hayes v Willoughby* [2013] UKSC 17, [2013] 1 WLR 935 at [1].

requirement is problematic. There is a strong risk of re-traumatising victims by asking them to provide evidence of the harm they experienced in order to secure a conviction. In addition, victims must experience distress before the perpetrator is at fault, which may not be the case – or may not be provable – despite a high level of culpability. The base offences of harassment and stalking under section 2 are also summary only offences, meaning the severity of the harm is not reflected in perpetrator culpability and, for the more serious offence under section 4 to apply, the harassment must cause the victim to fear that violence will be used against them.[83] While a victim of intimate image abuse will often experience fear, particularly where threats are made, this will not often be fear that physical violence will be used specifically. The more serious offence of stalking under section 4A includes where the course of conduct causes the victim to fear violence will be used against them *or* causes serious alarm or distress which has a substantial adverse effect on the victim's daily life but is again limited by the requirement for a course of conduct. Therefore, while some intimate image abuse may meet the required thresholds for these offences, much will not.

The communications offences under section 1 Malicious Communications Act 1988 and section 127 Communications Act 2003 could apply to intimate image abuse as they criminalise a range of grossly offensive, indecent, threatening and menacing communications. However, classifying conduct of this kind as a 'communications' offence does not fully convey the true nature and impact of the underlying offending behaviour. Section 127 is also a summary only offence, thus not reflecting the gravity of the harm caused by behaviour of this kind. Images must also be 'indecent' or 'grossly offensive' to engage the offence, which may well not be the case.

The parliamentary debate on the Policing and Crime Bill[84] noted that the offence of blackmail could be used for some forms of threats to disclose intimate images without consent, but only where threats were made with a view to make a gain or cause a loss.[85] Therefore, as noted in the Law Commission consultation, while this offence may cover some instances of threats to disclose, such as 'sextortion' where a victim is threatened with the sharing of their intimate image unless they pay money or send more intimate images, it is unlikely it could be utilised in the domestic context where threats are typically made to humiliate or distress, or to coerce and control the victim.

---

83 S 4(1).

84 Hansard, HL Deb, 16 November 2016, vol 776, col 1442.

85 While a 'gain' for these purposes need not be financial, the Court of Appeal has ruled that the gain must consist of 'property' (*R v Bevans (Ronald George Henry)* (1988) 87 Cr App R 64).

The offence of 'controlling or coercive behaviour in an intimate relationship' (the s 76 offence) was introduced in 2015 with the express aim of closing a legal 'loophole' by criminalising non-physical harm that was previously outside of the law's protection.[86] Notwithstanding the fact that this offence misconceptualises Stark's conceptualisation of coercive control, at first glance it appears that it would be the most appropriate charge where it is alleged that intimate images were shared, or threats to share were made, as a tool of coercion and control in an intimate relationship. Where the perpetrator has taken, made, shared, or threatened to share non-consensual intimate images, the requirements could be met if the behaviour is shown to be 'repeated or continuous' and has a 'serious effect' on the victim, 'serious alarm or distress which has a substantial adverse effect on B's usual day-to-day activities'.[87] Amendments recently introduced under the Domestic Abuse Act 2021 have also extended the offence of controlling or coercive behaviour under section 76 Serious Crime Act 2015 to former partners,[88] meaning that, where images are shared after the relationship ends, as a final act of 'revenge' or retaliation, or where threats are made as a tool of ongoing coercion and control, this could fall within the offence. There is no requirement that the prosecution proves the defendant was acting to cause a particular harm or for the purposes of sexual gratification, thus avoiding some of the limitations inherent in the offences discussed above, but there is a requirement that harm in the form of a 'serious effect' is established. Proof of harm was already shown to be a problematic requirement in the context of the harassment and stalking offences. In addition, for a successful prosecution it would need to be shown that the perpetrator (A) repeatedly or continuously engaged in behaviour towards the victim (B) that was controlling or coercive.[89] This could be problematic if there was just one (provable) threat or incident of sharing as it would not then be seen as repeated or continuous behaviour, thus limiting the scope of the offence in the same way as the requirement for a 'course of conduct' was shown to restrict the harassment and stalking

86 In introducing the clause on coercive control, then Attorney General Robert Buckland stated that the consultation 'identified a gap in the law – behaviour that we would regard as abuse that did not amount to violence. Violent behaviour already captured by the criminal law is outside the scope of the offence … we want an extra element that closes a loophole.' (Hansard, HC Deb, 20 January 2015, vol 591, col 172). As C Wiener points out in *Coercive Control and the Criminal Law* (Routledge forthcoming), this is empirically incorrect as coercive control consists of physical violence and other abusive tactics to gain and maintain power and control over the victim.

87 S 76(4).

88 S 68(4) Domestic Abuse Act 2021.

89 S 76(1)(a)

offences. As the discussion on separation assault and credible threats makes clear, one threat made to share an image if the victim leaves could well be enough to keep the victim entrapped in the relationship and subservient to the demands of the perpetrator. The dynamics of coercive control are hard to discern and can be hard to evidence due to the ways in which they can merge with socially accepted gender roles in heterosexual relationships.[90] Research also indicates that police still find it hard to identify coercive control, particularly where there is no evidence of serious physical violence.[91] These limitations mean that, while the s 76 offence may be used in this context, it is not likely to be applicable in many cases, and there are in fact indications that the offence is not being charged where intimate images are shared, or threats to share are made in the context of DVA.[92]

This analysis confirms the conclusions reached in both the *Shattering Lives* report and the Law Commission consultation; the existing patchwork of offences does not effectively criminalise all forms of intimate image abuse.[93] The protection offered is piecemeal and conceptually inconsistent, with many of the offences overlapping but using different language and terminology. The lack of coherence in the law leads to a 'failure to safeguard victims'[94] with different types of intimate image offences conceptualising the harm differently and requiring proof of different purposes and motivations. Different protections are also provided for victims depending on how the offence is classified. Reform is therefore essential to address the gaps and lacuna in the law and provide more effective protection for victims.[95] Four categories of intimate image offence are provisionally proposed on the basis that this would provide a more unified and structured approach by capturing the range of behaviours that constitute intimate image abuse, making the law governing this conduct clearer and more consistent, and ensuring that behaviour that is more culpable is dealt

90 See C Bishop and V 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 for a discussion of gender roles and evidencing coercive control.

91 A Robinson et al, 'Under the radar: policing non-violent domestic abuse in the US and UK' (2016) 40 International Journal of Comparative and Applied Criminal Justice 195; I Brennan and A Myhill, Domestic Abuse Matters 2.0: Evaluation of First Responder Training' (College of Policing 2017); Wiener (n 41).

92 Law Commission (n 1 above) para 3.160.

93 See ibid paras 3.130–3.201 for the Law Commission's analysis of these offences and their limitations.

94 Law Commission, Intimate Image Abuse: Summary of the Consultation Paper (Law Commission 2021).

95 Ibid.

with appropriately.[96] These will be outlined and then discussed in the next section.

LEGAL REFORMS

In the consultation, the Law Commission provisionally proposed the following offences:

1. a 'base' offence of taking or sharing an intimate image without consent, without a reasonable belief in consent, but with no additional intent element;
2. an 'additional' more serious offence of taking or sharing an intimate image without consent, with an intention to humiliate, alarm or distress the depicted person;
3. an 'additional' more serious offence of taking or sharing an intimate image without consent, without a reasonable belief in consent, for the purpose of obtaining sexual gratification, for oneself or another; and
4. an offence of threatening to share an intimate image.[97]

As well as seeking feedback on the proposed categories of offences, responses were also sought on a number of other issues including the proposed terminology and definitions. For the purposes of this article the focus will be on the consultation questions most pertinent to issues of coercion and control and the threats offence that was proposed, since these are of the most relevance where intimate images are used as tools of coercive control in abusive relationships.[98] The Law Commission classifies the conduct of perpetrators into three separate categories of taking, making and sharing an intimate image with the common thread being that 'the conduct takes place without the consent of the person in the image and violates their sexual privacy, autonomy and freedom,

96 Law Commission (n 1 above) para 14.8.

97 Shortly before publication of this article. the Law Commission published its final report (*Law Commission Intimate Image Abuse: A Final Report* (Law Com No 407, 2022)). This recommends five categories of intimate image offence - a "base" offence with no additional intent requirement, an installing offence, and the three more serious offences outlined above (at para 1.34).

98 There are 47 consultation questions in total, with the ones of specific interest to this article being those concerning the four offences that are provisionally proposed (consultation questions 26–28 and 40 at paras 10.60, 10.73, 10.79 and 12.138); whether proof of actual harm should be an element of intimate image offences (consultation question 24 at para 9.12.), whether an additional offence with a *mens rea* of 'intention to coerce or control' is needed (consultation question 30 at para 10.93); and whether a specific threats offence should be introduced (consultation question 40 at para 12.138).

their bodily privacy and their dignity'.[99] The consultation proposed that the intimate image abuse offences are classified as sexual offences, rather than as communications offences.[100] This is important for a number of reasons, including the availability of the protections and procedural safeguards that are already provided to victims of sexual offences, such as anonymity, special measures and limitations on cross-examination.[101]

Culpability, harm and the hierarchy of offences

In the consultation it is contended that proof of actual harm should not be an element of any new intimate image offences because it would be an unnecessary barrier to prosecution and could cause unnecessary distress to the victim.[102] This reflects the view of the majority of stakeholders during the pre-consultation stage, nearly all of whom agreed that offences of this kind should be categorised as criminal regardless of impact and that there should be no requirement for proof of harm.[103] Rackley and Johnson rightly point out that harm to the victim can be considered at the sentencing stage.[104] Critics of New Zealand's harmful digital communications offence[105] claim that the requirement that actual harm to the victim be proven makes the threshold for prosecution too high and whether a victim can prove that they were harmed is 'a subjective and arbitrary determination of whether an offence has occurred'.[106] An approach that requires proof of harm, as required under the s 76 offence, is often problematic and off-putting for victims because of the risk of re-traumatisation through an offence that requires them to give evidence of how and to what extent they were harmed by the actions of the defendant.[107]

Based on the difficulties inherent in the s 76 offence, it could be argued that there should be a 'proof of harm' requirement for any new

99 Law Commission (n 1 above) at para 1.15.

100 Ibid para 14.32.

101 The measures are outlined at n 73 above. Details of cases involving the disclosure of intimate images are widely reported in the press, suggesting the cases are going to cross-examination, underlining the importance of these protections in cases of this kind (A Dymock and C van der Westhuizen, 'A dish served cold: targeting revenge in revenge pornography' (2019) 39 Legal Studies 361, 370)

102 Law Commission (n 1 above) para 9.11.

103 Ibid para 9.6.

104 Ibid

105 S 22 Harmful Digital Communications Act 2015.

106 N Macdonald, 'Revenge porn: is the Harmful Digital Communications Act working?' (*Stuff*, 9 March 2019).

107 C Bishop cited in Law Commission (n 1 above) at para 9.8.

intimate image offence. However, it is suggested that this would only be necessary if there were to be just one basic offence, established where the defendant shared images without the consent of the victim, introduced, rather than a ladder of offences with different levels of culpability, similar to the non-fatal offences under the Offences Against the Person Act 1861 (OAPA 1861).[108] Whilst it is apparent that there are some benefits to the former approach as it would be straightforward to prove and pivots solely on the defendant's state of mind with regards to the victim's lack of consent (in the same way many sexual offences do), it is contended that this would be problematic because, without culpability, the harm inflicted on the victim may not be reflected in the penalties for the proposed offences.[109] Given that the framework of offences proposed by the Law Commission provides for different levels of culpability through the inclusion of a basic offence, two additional intent offences, and a threats offence, the need for a harm-based approach is unnecessary because culpability can be demonstrated through the other elements of the offence. However, it is vital that the final recommendations to Parliament clearly reflect the nature and extent of the harm of the different forms of intimate image abuse and establish the need for different levels of culpability.[110]

In England and Wales culpability is determined by reference to the level of harm caused and also by reference to the mental state of the offender at the time they caused the harm that constitutes the *actus reus* of the offence they are charged with.[111] This makes the

108 The concern that introducing a separate base offence and then more serious additional intent offences risks being overly complex for prosecutors, thus potentially impeding the effective prosecution of intimate image offences is reflected in the Law Commission consultation (n 1 above) at paras 10.94 and 10.95.

109 While the s 76 offence requires proof that the prohibited behaviour had a 'serious effect' on the victim, the offence and accompanying guidance does not demonstrate understanding of the nature of the harm to the victim. This is also seen in discussions prior to the introduction of the offence. This has resulted in a very low maximum sentence despite the serious harm that typically results from this type of behaviour (see Bishop and Bettinson (n 90 above)). The maximum prison sentence on indictment is 5 years (s 76 (11)(a)) with the maximum provided for in the Sentencing Council Guidelines being 4 years, with the presence of aggravating factors (Intimidatory Offences: Definitive Guidelines (Sentencing Council 2018)).

110 It is therefore to be welcomed that the final report, published shortly before this article was published, recommends a hierarchy of offences with different levels of culpability: Law Commission (n 97 above).

111 For example, under the OAPA 1861, culpability is predominantly approached in terms of physical harm (or, under *R v Chan Fook* [1994] 1 WLR 689, psychological harm where there is a recognised psychiatric condition) with offenders being held liable when they intentionally, or sometimes recklessly, cause harm to a victim.

outcome of an action, or the level of harm caused, an integral element when determining the severity of an offence and the corresponding penalties. However, it is not the only element; the fault element is concerned with both the defendant's state of mind at the time they committed the *actus reus* of the offence (the subjectivity principle) and the harm committed through their actions (correspondence principle).[112] This ensures a ladder of offences is constructed in 'ascending order of gravity'.[113] Although the fault element must relate to the harm committed, culpability is then determined by the defendant's state of mind.

This approach justifies the hierarchy of offences outlined in the consultation and negates concerns that including a number of offences with higher culpability carries the implication that *less* harm is caused where the defendant is liable only for the base offence. It is not that culpability is being increased based on an objective assessment of harm, thereby implying that the harm to the victim is less when the defendant acted without one of the intentions or purposes listed for the 'additional' offences. Instead, the defendant is deemed less *culpable* when not acting with one of the additional purposes or motivations and therefore a lesser penalty is justified. Indeed, it is contended that the offence itself reflects the level of harm, and then *culpability* is increased when the defendant acted with intention to cause a particular harm, rather than the additional offences reflecting greater harm to the victim.

The symbolic importance of this legislation cannot be underestimated. Attempting to implement fewer offences, or just one basic offence, to avoid impeding the prosecutorial process risks not adequately articulating the various types of harm or effectively reflecting the motivations and culpability of those who make, take and share intimate images. A ladder of offences is therefore needed to capture the various forms of intimate image abuse and the different motivations behind it. This would also emphasise the pervasive nature of intimate image abuse by bringing attention to the fact that it occurs in numerous contexts with various motivations and impacts, and therefore has a number of important implications for victims and wider society. It is not something that can be captured and dealt with under one simple offence. An offence with a very low maximum sentence, which would be the case with a low threshold for *mens rea* and no requirement for

112 In constructing offences, the Law Commission already recognises the importance of both the subjectivity and correspondence principles in determining defendant culpability. See Law Commission, *A New Homicide Act for England and Wales* (Law Com No 177, 2005) and also the ladder of non-fatal offences under the OAPA 1861.

113 Law Commission (n 1 above) para 1.32.

proof of harm to the victim, risks diluting the symbolic message about the severity of the consequences of intimate image abuse. This would fail to send a strong message about the serious nature of this crime and the severe, devastating and often lifelong implications it has for victims. The next subsection will consider how the motivations of the perpetrator where intimate images are utilised as tools of coercive control could be best accommodated within the proposed offences.

Motive: to control or coerce

The consultation invited views on whether there should be an additional offence of intentionally taking or sharing an intimate image without consent *with the intent to control or coerce* the person depicted, and whether this would be substantially different from an offence where the intent is to humiliate, alarm or distress the victim.[114] It is submitted that intention of this kind should be included within any new intimate image offences, either as a standalone offence or within the first additional offence (intention to distress, alarm or humiliate). An offence of this kind would more effectively encapsulate the nature of the harm inflicted on the victim and the motives of the defendant in certain situations, particularly where there is a state of coercive control maintained over the victim in an intimate relationship. This would reflect situations where images are shared by the perpetrator to restrict the victim's autonomy and freedom, prevent them from ending the relationship, and coerce them into behaving in a particular way. For example, images that have been taken, sometimes non-consensually, may be shared with a small group of people and this would leave the victim in fear of further images being taken and shared more widely. Indeed, it is known that the risk of future harm increases where intimate images exist and/or where threats have been made to share them.[115] Alternatively, where there was an intention to coerce or control it could be considered an aggravating factor at the sentencing stage, thereby increasing perpetrator culpability in that way.

During the pre-consultation some stakeholders suggested that where intimate image abuse takes place in the context of an intimate relationship it should be dealt with under existing domestic abuse legislation. However, while it is undoubtedly the case that where appropriate the s 76 offence should be used, and that police need to be trained to recognise the ways in which intimate images are used in a coercively controlling relationship, the s 76 offence, as outlined above, is far from satisfactory both in terms of the substantive law and its

114 Ibid para 10.92–10.93.

115 Citron and Franks (n 8 above); E Sharratt, 'Intimate image abuse in adults and under 18s' (South West Grid For Learning 2019).

implementation.[116] It should not be assumed that the s 76 offence can, or even should, be charged where intimate images are used as a tool of coercive control. Automatic anonymity for victims is also not granted for complainants where a s 76 offence is charged and therefore it may be preferable to charge under the proposed offences for this reason also, since automatic lifetime anonymity would be provided for victims under the legislation.[117] On the other hand, while naming coercion and control in this context is important for reasons outlined above, it may be that introducing too many new offences would be undesirable and overly complex, in which case it may be preferable to incorporate an intention to coerce or control within the second additional offence (taking or sharing an intimate image without consent, with an intention to humiliate, alarm or distress the depicted person) or as an aggravating factor in sentencing. A specific intention to 'coerce and control' could be incorporated in the context of threats to share.

Coercive control and the proposed threats offence

It is clear from the prevalence of threats to share intimate images, and the motivations of those who use them as a tool of coercive control, that the creation of a specific offence is justified.[118] As noted in the consultation, a number of threats offences already exist under criminal law, confirming that threats can warrant criminalisation even where they are not immediately actionable or effective, and that whether a threat has been made can be established objectively.[119] In recognition of the fact that threats to share intimate images are the most common type of threat, with evidence suggesting that these mainly take place in the context of abusive relationships,[120] the consultation proposed that it should be an offence for D to threaten to share an intimate image of V, where:

116 Domestic Abuse and the Criminal Justice System, England and Wales: November 2020 (Office for National Statistics 2020); *Everyone's Business: Improving the Police Response to Domestic Abuse* (HM Inspectorate of Constabulary 2014); J Youngs, 'Domestic violence and the criminal law: reconceptualising reform' (2015) 79 Journal of Criminal Law 55; Bettinson and Bishop (n 81 above); C 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 2016); Bishop and Bettinson (n 90 above); M Hester, 'Making it through the criminal justice system: attrition and domestic violence' (2006) 5 Social Policy and Society 79; Wiener (n 41); Brennan and Myhill (n 91 above).

117 Law Commission (n 1 above) para 14.85.

118 As suggested by Refuge, criminalising threats could mean images are never actually shared, something that currently happens in 23% of cases where a threat has been made: Refuge (n 7 above).

119 Law Commission (n 1 above) 12.24

120 Ibid para 12.3

(a) D intends to cause V to fear that the threat will be carried out; or

(b) D is reckless as to whether V will fear that the threat will be carried out.[121]

In proposing the introduction of a specific threats offence, the Law Commission took the lead from jurisdictions such as Australia, where threats constitute a separate offence[122] rather than following in Scotland's footsteps where threats are combined with the taking, making and sharing offences.[123] In doing so the different character of threats as compared with taking, making and sharing is noted; here the harm arises from the threat itself rather than the taking, making or sharing that may or may not follow. Creating a specific threats offence means that it can be tailored to ensure that only harmful behaviour is criminalised and that the elements are not unduly restricted by a focus on the taking, making or sharing offences.[124] In the consultation the Law Commission submitted that an alternative additional intent requirement, such as the intent to cause distress, is unnecessary for the threats offence due to the proposed fault element. The argument behind this is that, '[i]nherent in a threat which the defendant intends the victim to fear will be carried out (or is reckless thereto) is an intention

121 Ibid para 12.138. The Law Commission proposes that the same definition of 'intimate image' is used for both the offences of sharing and threatening to share an intimate image (para 12.139). The test of recklessness would be the same as for other criminal offences, that is to say a subjective test with an objective element (*R v Cunningham* [1957] 2 QB 396 as confirmed by *R v G and R* [2004] 1 AC 1034). Although the majority of consultees who responded to this question supported an additional offence with an intent to control or coerce the person depicted, the Law Commission ultimately did not recommend an offence of this kind. This decision was made on the basis that existing offences and the other intimate image offences satisfactorily cover a large range of culpable intimate image abuse conducted to control or coerce the person depicted (Law Commission (n 97 above) para 6.158). However, it did recommend that the Government consider reviewing the statutory guidance for the offence of controlling or coercive behaviour in light of the ways in which intimate image abuse is perpetrated in the context of abusive relationships (ibid para 6.166)

122 There are separate threats offences in Victoria under s 41DB of the Summary Offences Act 1966, in New South Wales by virtue of s 91R Crimes Amendment (Intimate Images) Act 2017 No 29 (NSW), in Western Australia where the Criminal Law Amendment (Intimate Images) Act 2019 amended the Criminal Code to make an offence of distributing an intimate image of another person without their consent, or threatening to distribute an intimate image of another, and in the Australian Capital Territory where the Crimes (Intimate Image Abuse) Amendment Act 2017 amended the Crimes Act 1900 to include intimate image offences, with s 72E of the Crimes Act 1900 creating a specific threats offence.

123 S 2 Abusive Behaviour and Sexual Harm (Scotland) Act 2016.

124 Law Commission (n 1 above) at para 12.114.

to cause the victim distress (or recklessness as to whether distress is caused)', making such an additional intent requirement superfluous.[125] However, there may be a different offence to articulate here, where D threatens to share an image with the intention of controlling or coercing V. This seems to make D more culpable than where they intentionally or recklessly threaten to share an image without comprehending how much harm might be caused to V through this apprehension (ie where they intend or are reckless as to causing V to fear the image may be shared, but do not consider the impact that this fear may have on V). The impact of a threat to distribute intimate images extends beyond the humiliation or shame that a survivor might be concerned about should the abuser follow through with the threat.[126] It is important that the harmful impact of the behaviour and the motivations of the perpetrator are articulated and reflected in the offence.

Of course, where threats to share are made as part of a clear and provable programme of coercive control the s 76 offence is available and should be used. However, as discussed above, there may be situations in which it is not possible to evidence this offence or it is not relevant or desirable to bring a charge for other reasons, and therefore it is important that the proposed threats offence can accommodate the more serious nature of the harm and culpability where threats form part of an overall programme of coercive control. An objection may be raised that having two threats offences would lead to the creation of too many offences in total. However, this implies that threats to share are less serious than situations where the images are in fact shared, something which the above discussion shows is often not the case. It is this author's view that the threats offence should articulate the intentional or reckless use of threats in order to control or coerce the victim as a different harm and level of culpability than where the defendant threatens to share images solely intending or being reckless that V will fear the image will be shared.[127]

The Law Commission expresses understandable reluctance to criminalise the mere possession of intimate images without consent to avoid over-criminalisation.[128] However, in the context of DVA, particularly where there is coercive control, the existence of an image

125 Ibid para 12.133.

126 Cuomo and Dolci (n 10 above) 230.

127 However, in light of the clear evidence that intimate image abuse is often a part of controlling or coercive behaviour put forward by this author and other consultees, it did recommend that the Sentencing Council should consider whether an intent to control or coerce should be an aggravating factor at sentencing for the offence of threatening to share an intimate image (Law Commission (n 97 above) para 12.163). This therefore reflects the argument put forward above that conduct of this kind is more culpable and therefore warrants a higher penalty.

128 Law Commission (n 1 above) consultation question 18 at para 7.86.

can harm the victim even where there is no explicit threat to share it. It may prevent them from leaving or attempting to leave and deter them from disclosing the violence and abuse, reporting to the police, or seeking help because they are aware that the images exist and that they could be used against them. It is for this reason that threats – for the purposes of the proposed threats offence – must be broadly construed to include implicit as well as explicit threats. This would reflect the New South Wales threats offence which makes clear that the threats offence can be committed by any conduct, explicit or implicit, conditional or unconditional.[129] Training and guidance around this will need to be thorough and inclusive, and where the *existence* of images, without threats to share, is responsible for keeping the victim in an abusive relationship, or stopping them reporting, disclosing or seeking help, then the s 76 offence should be used where relevant.

The preceding discussion emphasises the importance of considering the introduction of an additional offence of 'intentionally taking or sharing an intimate image without consent with the intent to control or coerce the person'. However, including an intention to coerce and control as an aggravating factor at the sentencing stage could also be considered. The creation of a specific offence of threatening to share has also been justified, with the suggestion that incorporating an intention to coerce and control in this context may better reflect the culpability of those who engage in *threats* to share intimate images and should therefore be introduced not just where images are taken, made and shared, but also where threats to share such images are made.[130] It is suggested that, were these offences to form part of the framework of reforms implemented, it would significantly increase the ability of the criminal law to protect victims and impose penalties befitting the severity and nature of the harm intended and caused.

CONCLUSION

The Law Commission's current project on taking, making and sharing intimate images without consent provides an important opportunity to address one of the most pervasive and ubiquitous tools utilised by perpetrators to maintain power and control in intimate relationships. In highlighting the role of intimate image abuse as a tool of coercive control, this article has emphasised the importance of the law fully recognising the harm that it seeks to remedy and the motivations behind the behaviour that causes this harm. It is therefore clear that the harmful loss of autonomy and personhood that results from the use

129 91R Crimes Amendment (Intimate Images) Act 2017 No 29 (NSW).

130 Law Commission (n 1 above) para 10.93.

of intimate images as tools of coercive control must be reflected both in existing offences and in any potential new offences and must be used to determine both the level of culpability and appropriate penalties. Detailed consideration of the offences proposed in the consultation confirms that there is the potential for legal reforms to effectively conceptualise intimate image abuse as a tool of coercion and control and to reflect the motivations of those who engage in this behaviour through the inclusion of several offences with differing levels of culpability.

Of course, substantive law reform can only ever be one strategy used to address gendered harms and violence against women and girls. Victim reluctance to report or participate in the criminal justice process, alongside the widely documented barriers to successful conviction of gender-based crimes, make it clear that introducing new legislation will not be a panacea when it comes to increasing protection and bringing perpetrators to justice.[131] In addition, intimate image abuse, particularly when perpetrated against partners and former partners, sustains and normalises male power and privilege in the domestic sphere at a cultural level. This highlights that the taking, making and sharing of intimate images without consent is harmful not only to individual victims but also to women as a group, due to the normalisation of harmful sex-role stereotypes it facilitates, indicating that wider social change, that goes beyond the criminalisation of individual perpetrators, is needed to fully address this issue. However, criminal law is an important first step in facilitating the wider changes needed as it can play an important role in shaping these new manners or ethics

131 Unwillingness to report, retraction of complaints and the withdrawal of support for prosecution is already an issue with DVA, sexual offences and prosecutions under the 'disclosure offence'. A BBC freedom of information request revealed that no action was taken in 61% of cases of 'revenge porn' recorded by the police between April and December 2015. Many of these reported incidents failed to proceed due to evidential difficulties or because the complainant had withdrawn their support for prosecution (P Sherlock, 'Revenge pornography victims as young as 11, investigation finds' (*BBC News* 27 April 2016)). For attrition in domestic violence cases, see A Cretney and G Davis, 'Prosecuting domestic assault: victims failing courts, or courts failing victims?' (1997) 36 Howard Journal of Criminal Justice 146; L Bennett et al, 'Systemic obstacles to the criminal prosecution of a battering partner: a victim perspective' (1999) 14 Journal of Interpersonal Violence 761; Hester (n 116 above); A Robinson and D Cook, 'Understanding victim retraction in cases of domestic violence: specialist courts, government policy, and victim-centred justice' (2006) 9 Contemporary Justice Review 189. With regards rape, it is estimated that only 15% of victims report to the police (Ministry of Justice, Home Office and the Office for National Statistics, *An Overview of Sexual Offending in England and Wales* (Home Office January 2013)) and an extremely high attrition rate means only 1.4% of reported rapes result in even a charge or summary (D Shaw, 'Rape convictions fall to record low in England and Wales' (*BBC News* 30 July 2020)).

and new offences can strengthen perceptions about the wrongfulness of behaviour.[132] This is particularly relevant in the context of intimate image abuse because it is a – relatively – new behaviour and therefore new norms are still forming.[133] Legal reforms that condemn and appropriately punish harmful behaviours are therefore an important first step towards the wider change needed at the same time as being able to ensure that the harmful loss of autonomy resulting from the sharing of images, or threats to share images, in the context of DVA is appropriately captured in the legal framework.

132 Crofts and Kirchengast (n 8 above) 93. See also D Citron, 'Sexual privacy' (2019) 128 Yale Law Journal 1874; Citron and Franks (n 8 above); D Citron and J Penney, 'When law frees us to speak' (2019) 87 Fordham Law Review 1; McGlynn and Rackley (n 8 above).

133 Crofts and Kirchengast (n 8 above).

Northern Ireland
Legal Quarterly
Autumn Vol. 74.3 (2023) 588–618
Articles DOI: 10.53386/nilq.v74i3.1121

International law for public health in aviation: the challenges of harmonisation†

Elizabeth M Speakman
University of Dundee
Mary Cameron
London School of Hygiene and Tropical Medicine
Andrea Grout
James Cook University, Australia
Correspondence email: espeakman001@dundee.ac.uk

ABSTRACT

International laws for commercial aviation have achieved an exceptional degree of harmonisation and greatly improved passenger safety. Yet, despite much international guidance, enforceable laws for public health protection in aviation are mainly the responsibility of national authorities. As a result, public health laws may be incoherent, in conflict with other countries and/or based on disputed scientific evidence. The COVID-19 pandemic has highlighted the responsibility of airlines and regulatory authorities to protect not only air passengers but also populations in destination countries. While the greatest risk to global public health is the potential spread of disease by infected passengers or vectors, lesser-known risks include food contamination, inadequate sanitary facilities and poor air quality within the cabin. In preparedness for inevitable future disease outbreaks and pandemics, an urgent review of international law as it applies to public health in commercial aviation is needed, with greater investment in scientific research to enable more accurate and effective risk assessment and management, supported by enforceable laws and clear responsibility.

Keywords: aviation; public health; law; harmonisation.

INTRODUCTION

The global impact of COVID-19 and the role of commercial aviation in its rapid worldwide spread mandates an urgent re-examination of public health risks in commercial air travel, including the limitations of international aviation laws and regulations in managing those risks. Aviation law is a well-established field of legal expertise, and international public health law has been explored by scholars for decades, but there is a surprising lack of academic literature on international public health law as it applies to commercial aviation.

† First published in NILQ 73.AD1 130–160 27 July 2022.

In 2019 Cuinn and Switzer[1] examined the fragmentary nature of international aviation governance during the 2014 Ebola outbreak in West Africa, but there are very few other examples.

There remain important questions about the duty of care of international commercial enterprises for public safety, not just to their paying customers but to the global population. Balancing the conflicting pressures of public safety and commerce can be exceptionally difficult in cases where science does not yet have clear answers. Apart from the devastating human and social cost, the pandemic has inflicted unprecedented economic damage on the aviation industry, with predicted lost revenue of $314 billion for 2020, a drop of 55 per cent from 2019.[2] Even with financial support from national governments and radical cost-cutting measures (including mass redundancies), many airlines may not recover from the economic impact.[3] Dube noted the low resilience of the industry and that early signs indicate a 'slow, unpredictable and stretched recovery' with 'colossal cash burn'.[4] Gössling argued that the conflict between volume growth and risks and vulnerabilities is insurmountable, and that the commercial aviation industry should be deliberately shrunk, with financial subsidies reduced or withdrawn.[5] It is clear that commercial aviation faces considerable challenges ahead.

The present article provides an overview of the current international legal framework for public health protection in aviation, some of the key public health risks in commercial air travel with particular focus on risks within the aircraft cabin and provides examples of governance to manage those risks. It looks at liabilities and considers the challenges of developing harmonised, enforceable legislation for public health protection. It does not attempt to give a comprehensive account of all international aviation law, nor can it address the many different national legislative frameworks around the world. It argues that the damage to the aviation industry and its threat to global health requires a 'reset'. There is an urgent need for industry and public health leaders to collaborate in a comprehensive assessment of public health risks

1 Gearóid Ó Cuinn, and Stephanie Switzer, 'Ebola and the airplane – securing mobility through regime interactions and legal adaptation' (2019) 32(1) Leiden Journal of International Law 71–89.

2 IATA Press Release No 29, 'COVID-19 puts over half of 2020 passenger revenues at risk' (14 April 2020).

3 'COVID 19: "Future of UK aviation" at risk, say airlines' (*BBC News* 15 March 2020).

4 Kaitano Dube, Godwell Nhamo and David Chikodzi, 'COVID-19 pandemic and prospects for recovery of the global aviation industry' (2021) 92 Journal of Air Transport Management.

5 Stefan Gössling, 'Risks, resilience, and pathways to sustainable aviation: a COVID-19 Perspective' (2020) 89 Journal of Air Transport Management.

in aviation, and to identify where research investment is needed, how risks should be managed and who should take ultimate responsibility for enforcement.

OVERVIEW OF INTERNATIONAL LEGAL FRAMEWORK FOR AVIATION

'Legal framework' as used here refers to the broad system of rules that govern and regulate decision-making, agreements and laws, also known as 'governance'. Governance can include both *law* and *policy* and there are important differences between the two. Laws which have been enacted by a national government, by judicial precedent or by custom are regarded as 'hard law', that is, they are usually binding and enforceable. They are fixed and publicly available. 'Soft law' can include policy, guidelines and recommended practice which may have been created for an organisation or industry's internal procedures, and therefore may not be public. It is non-binding and non-enforceable but may carry influence. A particular advantage of soft law, and the reason it is common in international law, is that it can mobilise the consent of countries with different interests and where the commercial interests of private bodies are involved.[6] A key weakness is the lack of an enforcement mechanism, but nevertheless, the intention is to create 'norms' of behaviour, that is, it is 'normative' and by signing international treaties and conventions, countries commit to abide by the terms of those agreements. Sekalala and Masud argue that soft law may be both a precursor and complementary to hard law,[7] and it will be strengthened if its terms are incorporated and enacted in national laws, becoming enforceable.

The uniqueness of international aviation law lies in the level of state compliance with international treaties and how this has enabled harmonisation of national laws worldwide. The importance of internationally agreed rules was recognised from the earliest days of aviation and, despite fundamentally different political, legal and economic contexts, most nations have come together to commit to binding technical standards in aviation. While this was initially to protect air sovereignty, the regulations soon focused on security and the technical safety of aircraft to reduce accidents. The twentieth century saw a succession of major treaties which achieved an impressive degree of conformity and collaboration, such as the Warsaw Convention 1929, the Chicago Convention 1944 and the Montreal Convention

6 Sharifah Sekalala and Haleema Masud, 'Soft law possibilities in global health law' (2021) 49(1) Journal of Law, Medicine and Ethics 152–155.

7 Ibid.

1999, as well as the establishment of the International Civil Aviation Organization (ICAO).

Aviation law has become highly standardised, as well as increasingly complex and specialised, including many subspecialties such as access to airspace, contractual and commercial law, environmental law and now expanding to include unmanned drones and space law. Yet there is a lack of harmonised, enforceable international law for *public health risks* in aviation. International public health is the domain of the World Health Organization (WHO), and WHO's International Health Regulations (2005) (IHR) have specific annexes for aircraft, but these and other international guidelines are unenforceable. National laws to reduce public health risks in aviation may be limited, in conflict with those of other countries or based on still evolving (and disputed) scientific evidence. Public health threats include on-board food contamination, inadequate facilities and poor air quality (although the last is fiercely disputed by the aviation industry). The greatest risk is of spreading highly virulent diseases by carriage of infected passengers or vectors. This is primarily through the movement of infected individuals to new geographical locations, rather than transmission occurring on board aircraft. While rare, such events can be catastrophic and endanger populations.

CURRENT INTERNATIONAL LEGAL FRAMEWORK FOR AVIATION: INTERNATIONAL REGULATORY BODIES

While WHO is the body with overall authority for global public health, including in aviation, the most important regulatory body specifically for aviation is ICAO, an official body and specialised agency of the United Nations (UN) which was established by the Chicago Convention in 1944. All 193 current member states of the ICAO have committed to accept ICAO Standards, and oversight and enforcement of the regulations is usually the responsibility of the National Civil Aviation Authorities (NCAAs) of each country. Thus a required ICAO Standard for a particular technical modification must be enacted in all 193 member states and be enforceable in each country under national laws.

ICAO's core mandate is 'to help States to achieve the highest possible degree of uniformity in civil aviation regulations, standards, procedures and organization'.[8] Over decades and by consensus of its members, ICAO developed 19 annexes containing over 12,000 Standards and Recommended Practices (SARPs) and five Procedures for Air Navigation (PANs) concerning mechanical safety, aircrew training, use of commercial airspace, environmental controls and

8 ICAO, 'The history of ICAO and the Chicago Convention'.

many more. The SARPs focus on issues such as mechanical safety of the aircraft, aircrew qualifications, right to airspace, customs and freight and air traffic control. In recent years aircraft emissions have also become an area of regulation, as well as working conditions of aircrew and measures to prevent aviation terrorism. 'Standards' are technical specifications, 'the uniform application of which is recognised as necessary for the safety or regularity of international air navigation and to which Contracting States will conform in accordance with the Convention'. Recommended practices are deemed 'desirable in the interest of safety, regularity or efficiency of international air navigation and to which Contracting States will endeavour to conform in accordance with the Convention'. Thus, standards are considered obligatory, while recommended practices are advisory. States may still avoid compliance with standards if they file a 'difference' with ICAO although this may result in penalties. For example, another state may prevent aircraft with these reduced requirements from entering its own airspace. Also, any state can apply higher national standards than those of ICAO without penalty.

While the system is imperfect, with evidence of some countries falling behind in compliance,[9] in general SARPs have contributed to enormous progress in improving mechanical safety, upheld by international and domestic law. The 19 current annexes contain references to public health issues, but these are limited, and in most cases simply require compliance with WHO guidelines.

The founding of ICAO was rapidly followed in 1945 by the establishment of the International Air Transport Association (IATA), a trade association which has grown to a current membership of 290 airlines from 120 countries,[10] accounting for 83 per cent of total air traffic.[11] IATA has issued many important safety guidelines and valuable guidance through the work of its medical advisor and Medical Advisory Group. However, it carries less authority than ICAO, its standards and guidelines are non-binding and, as a trade association, its priority is the interests of the airline industry.

Other important organisations in international aviation include Airports Council International (ACI) and the International Flight Services Association (IFSA). ACI is a membership body which represents airports across the world. It promotes cooperation between airports and often works with other regulatory bodies, primarily ICAO and IATA, as well as developing its own standards, recommended practices and policies for safety and security. Its role is to 'represent

9 ICAO Safety Audit Results.
10 IATA Current Airline Members.
11 IATA, 'About us'.

the collective interests of airports around the world'.[12] As of January 2022, it had 701 members operating 1933 airports in 183 countries.[13] Apart from its collaborative regulatory work with other bodies it has produced its own *Policy Handbook*.[14] IFSA is a global professional association which was created in 1966 'to serve the needs and interests of airline and railway personnel, inflight and rail caterers and suppliers responsible for providing onboard services on regularly scheduled travel routes'.[15] It has a particular focus on onboard food safety.

REGIONAL REGULATORY BODIES

There are a number of regulatory bodies based in Europe which work to harmonise European aviation governance and to support ICAO. The European Civil Aviation Conference (ECAC) was founded in 1955 with a mission to promote 'the continued development of a safe, efficient and sustainable European air transport system'.[16] As an intergovernmental organisation of 44 European member states, ECAC issues guidelines, policy recommendations and resolutions. The European Organisation for the Safety of Air Navigation (EUROCONTROL) was founded in 1960 and provides technical expertise in relation to air traffic management across Europe.[17]

The European Union (EU) established a single aviation market for Europe in 1992. This European 'open skies policy' is probably unique in the world and has provided much commercial benefit, particularly for low-cost European airlines. A development from this cross-border integration was the establishment of Joint Aviation Authorities (JAA) which evolved into the European Aviation Safety Agency (EASA) in 2002.[18] As with NCAAs, EASA can recommend changes to existing regulations or the introduction of new regulations, but these are enacted by the EU and EASA's role is in oversight and enforcement. EASA is becoming an increasingly important regulatory actor in the European region. Notable EU legislation has included Regulation 2111/2005 which introduced a 'blacklist' of carriers banned from

12 ACI, 'About ACI'.
13 ACI website.
14 ACI, *ACI Policy Handbook* 9th edn (ACI 2018) i–ii.
15 IFSA, 'About'.
16 ECAC, 'About ECAC'.
17 EUROCONTROL website.
18 Regulation (EC) No 1592/2002 of 15 July 2002 on common rules in the field of civil aviation and establishing a European Aviation Safety Agency.

operating within EU air space.[19] Further conformity within EU air space came with Regulation (EC) No 216/2008 of 20 February 2008 on common rules in the field of civil aviation. EASA explicitly aims for compatibility with ICAO annex 19 on safety management.[20] EU regulations are enforceable within member states, making EASA an important source of international aviation law.

In Asia, the area of fastest growth in aviation, the Single Aviation Market (SAM) was established in 2015, an initiative of the Association of South-East Asian Nations (ASEAN). At present this is a commercial arrangement and any attempt to introduce harmonised public health standards in aviation for Asia would be very challenging given the diverse national contexts within ASEAN and the enormous size of the Asian aviation market.

Much aviation governance is created by consensus of these groups. However, only ICAO and EASA have legislative power over member states: ICAO at international level, EASA at regional level. Even for ICAO and EASA there are limits of enforceability, which usually takes the form of penalisation of members.

Mention should also be made of the Federal Aviation Administration (FAA) which is an agency within the United States (US) Department of Transportation and the regulatory authority for aviation in the US, replacing the Civil Aeronautics Administration (CAA) in 1958. FAA rules are set out in the Federal Aviation Regulations (FARs), also known as Title 14 Code of Federal Regulations (CFR), which are binding and enforceable and are intended to ensure aviation safety in the US. Although a national agency, the positions taken by FAA on regulatory issues are highly influential but have also received repeated criticism for allegedly being too heavily influenced by the US airline industry.[21, 22]

---

19 Regulation (EC) No 2111/2005 of the European Parliament and of the Council of 14 December 2005 on the establishment of a Community list of air carriers subject to an operating ban within the Community and on informing air transport passengers of the identity of the operating air carrier, and repealing Article 9 of Directive 2004/36/EC.

20 EASA, 'SMS – EASA Rules'.

21 David B Carmichael, Mary N Kutz and Dovie M Brown, 'FAA "captured?" is the Federal Aviation Administration subject to "capture" by the aviation industry?' (2003) 21(1) Collegiate Aviation Review International.

22 Stephen Mihm, 'The FAA has always been cozy with the aviation industry. That's why we need to empower the NTSB' *Los Angeles Times* (22 March 2019).

CURRENT INTERNATIONAL LEGAL FRAMEWORK: KEY GOVERNANCE FOR PUBLIC HEALTH PROTECTION IN AVIATION

The IHR (2005) are the pre-eminent legal instrument for global public health, and 196 sovereign states[23] have committed to be legally bound by their terms. At the time of their creation, Fidler argued that the IHR represented 'a significant shift in international health cooperation',[24] which represented 'a conceptual breakthrough in global governance. Instead of commercial interests defining the scope and purpose of the IHR, public health considerations now take priority.'[25] Some years later, Gostin noted that:

> Finding ways to balance public health and economic activity has become an enduring feature of global governance ... The revised IHR sought to promote greater state compliance. Yet the regulations grant the WHO few, if any, explicit powers to monitor state performance, impose sanctions, or provide incentives ... Instead, the IHR rely on global norms and transparency, as civil society and the international community hold states accountable for evidence-based decisions.[26]

While intended to be binding, as Gostin points out, there is no enforcement mechanism to the IHR, and it is therefore soft law. Member states retain the sovereign right to legislate in accordance with their own health policies, but are expected to uphold the regulations[27] and much regional and national public health regulation incorporates its terms either implicitly or explicitly. The IHR include a number of terms which apply to air transport and aviation, including part IV – 'Points of entry' – and part V – 'Public health measures'. For example, part V, chapter II, article 24, 1(c) requires states to 'take all practicable measures' to 'permanently keep conveyances for which they are responsible free of sources of infection or contamination, including vectors and reservoirs. The application of measures to control sources of infection or contamination may be required if evidence is found.'[28] More specific provisions are set out in annexes 4 and 5 of the IHR, and the WHO regularly produces guidance developed by world experts in public health, and generally in collaboration with the aviation industry. The IHR are a critical foundation of public health in aviation, but their terms are general and unenforceable due to the sovereignty

23 194 WHO member states, Liechtenstein and the Holy See.

24 David Fidler, 'From international sanitary conventions to global health security: the new International Health Regulations' (2005) 4(2) Chinese Journal of International Law 325–392.

25 Ibid.

26 Lawrence Gostin, *Global Health Law* (Harvard University Press 2014) 183, 197.

27 IHR (2005), pt II, *Art 3 Principles*, para 4.

28 Ibid, pt V, ch II, art 24 'Conveyance operators', 1(c).

of member states. A recent report by WHO on the functioning of the IHR during the current COVID pandemic found that 'in the context of a pandemic, countries that in 2005 approved the IHR, in 2020 only applied the Regulations in part, were not sufficiently aware of them, or deliberately ignored them'.[29]

Apart from the IHR, the most significant international treaties specifically for health protection in aviation are the Warsaw Convention 1929,[30] the Chicago Convention 1944[31] and the Montreal Convention 1999.[32] The Warsaw Convention was important for being the first international agreement which imposed a strict (if limited) liability on commercial airlines for any event causing injury or death to passengers. The Chicago Convention was transformative for its establishment of ICAO at a critical time politically and in terms of the technological development of aircraft. The Montreal Convention in 1999 largely replaced the Warsaw Convention in its expansion of rights for passengers.

ICAO regulations (19 annexes to the Convention) cover a broad spectrum of aviation safety issues. For example, annex 1 concerns personnel licensing, and annex 8, 'Airworthiness of aircraft', is specifically concerned with mechanical safety. Yet, compared to the precise, binding laws in annexes 1 and 8, regulations for public health protection in aviation are permissive rather than mandatory.

Annexes which might have public health implications are annex 6 ('Operation of aircraft'), annex 9 ('Facilitation'), annex 11 ('Air traffic services') and annex 14 ('Aerodromes'). The first edition of annex 9 ('Facilitation'), published in 1953, includes chapter 8 ('Sanitation, medical services and agricultural quarantine') and says that contracted states should comply with the provisions of the International Sanitary Regulations (WHO Regulations No 2), accept WHO International Certificates of Vaccination and Revaccination, and should accept public health information in the form provided in the General Declaration. However, this is a recommendation, not a standard. ICAO has been transferring common safety elements from these annexes to a new annex 19 for 'Safety management'. The SARPs in annex 19 'shall be

29 WHO, 'Report of the Review Committee on the Functioning of the International Health Regulations (2005) during the COVID-19 response' (WHO 30 April 2021) 7.

30 Convention for the Unification of Certain Rules Relating to International Transport by Air 1929.

31 Convention on International Civil Aviation 1944.

32 Convention on the Unification of Certain Rules for International Carriage by Air 1999.

applicable to safety management functions related to, or in direct support of, the safe operation of aircraft'.[33]

The SARS outbreak of 2003 led to increased cooperation between IATA, ICAO and WHO.[34] This collaboration later included ACI and has been vitally important during major incidents. But there is far less clarity on the best response to a range of less visible public health risks on international flights, or identifying which authority has ultimate responsibility for coordinating this response. While WHO is the coordinating body for responses to Public Health Emergencies of International Concern (PHEICs), with its mandate based on the IHR (2005), its remit is the protection of *global* public health. The commercial aviation industry, through ICAO and other member bodies, has instead focused on *passenger* safety. WHO has only limited oversight regarding air transportation and already faces major challenges relating to the effectiveness of the IHR because of the need to support weak health systems in low-income countries. Thus, although WHO coordinates with aviation and state regulators, a greater rule-making role in civil aviation would almost certainly be beyond WHO's capacity, mandate or acceptability to stakeholders.

PUBLIC HEALTH RISKS IN AVIATION

As aviation technology developed throughout the twentieth century, the focus was on achieving mechanical safety. Since the early days of commercial flights, accidents have become rare events – testament to the success of these measures. Recent years have witnessed huge changes in commercial aviation, with rapidly increasing passenger numbers, longer flights and extended flight networks to previously isolated regions. The current aviation industry, with a multitude of short-haul low-cost airlines, and at the other extreme, ultra-long-haul flights (lasting 16 hours or longer),[35] would be unrecognisable to the early aviation pioneers, or even to the delegates at the Chicago Convention in 1944. Prior to the COVID-19 outbreak, the commercial aviation industry was projected to expand rapidly in the coming decades, with the fastest growth in Asia and developing countries. Annual international passenger numbers stood at 1.467 billion in 1998, had grown to 3.979 billion by 2017 and with an annual growth rate of

33 ICAO, Annex 19 to the Convention on International Civil Aviation, Safety Management, 2nd edn, July 2016, ch 2 'Applicability'.

34 Cuinn and Switzer (n 1 above).

35 Ultra-long range operations (ULRs) are 'flight operations involving any sector between a specific city pair in which the planned flight time exceeds 16 hours, taking into account mean wind conditions and seasonal changes' (ICAO 2012).

3.5 per cent were forecast to reach 8.2 billion by 2037.[36] More fuel-efficient aircraft, low oil prices and customer demand made ultra-long-haul flights more common.[37] Whether this growth trajectory recovers post-COVID-19, or alternatively, the industry suffers long-term loss of public confidence, the need for robust, evidence-based, yet adaptable regulatory mechanisms is greater than ever.

The constant transport of large numbers of people across the globe brings public health risks for passengers, aircrew and the populations in destination countries. Prevention and response to public health threats require different, sometimes highly complex measures. For many of these threats there are inadequate data (partly due to a lack of monitoring) and no scientific consensus. While an impressive global conformity was achieved in technical safety standards, the same cannot be said for public health protection. Although the governance bodies described above have frequently collaborated with each other to develop guidelines, including sections of the IHR (2005), these carry less weight than the SARPS and are unenforceable. With the notable exception of environmental regulations, to date, neither ICAO nor EASA has introduced any binding international law for public health protection. Many individual countries have introduced relevant national legislation but, since these are not internationally harmonised and are sometimes based on differing scientific evidence, they may result in conflict of laws.

Aircrew are trained to be first responders and all commercial flights should carry a supply of emergency medical kit. Medically trained passengers are often asked to help out and there is also increasing reliance on medical advisors on the ground. However, there is no universally agreed kit and legal requirements vary across countries. In 2016 the FAA granted exemptions to 50 airlines from carrying a range of emergency medications.[38] Furthermore, a comparative study of American, European, Indian, Indonesian, Emirati and Canadian civil aviation regulations for carriage of first-aid and emergency medical kits found a lack of transparency, variation in criteria and exemptions.[39]

---

36 IATA Press Release No 62, 'IATA forecast predicts 8.2 billion air travelers in 2037' (24 October 2018).

37 Graphic Detail, 'The rise of the ultra-long-haul flight' (*The Economist* 27 March 2018).

38 Federal Aviation Administration Exemption Number: 10690E 29 January 2016.

39 Wilfredo Rodriguez-Jimenez, 'First aid kit and emergency medical kit onboard commercial aircraft: a comparative study of American, European, Indian, Indonesian, Emirati and Canadian Civil Aviation Regulations' (MPH, University of Texas Medical Branch 2017).

Vulnerable passengers

Passenger demographics have changed, with increased travel by the elderly, disabled and those with chronic illnesses.[40] Silverman and Gendreau[41] noted how passengers differ in vulnerability, and the pool of highly susceptible individuals is likely to increase. The UK Government reported that requests for special assistance at UK airports 'are increasing at a rate of around double that of general growth in passenger numbers'.[42] While not a direct public health threat in itself, this may create a greater potential for inflight incidents related to a susceptible condition. It may also mean an increase in workload for cabin crew. With any illness or medical condition, the risk of an inflight medical emergency increases, which in turn can impact flight safety (eg by diverting to alternate airports). The majority of inflight emergencies were due to exacerbation of pre-existing medical problems (65 per cent)[43] and ultra-long-haul flights put particular stress on such passengers. Syncope (temporary loss of consciousness) is the most common inflight medical emergency, accounting for 91 per cent of new inflight emergencies, and is considered likely related to a prolonged period of sitting.[44] The rarity of syncope during long-distance bus or rail travel suggests that air cabin pressure or air quality might be contributory factors.[45] A greater distance travelled is a significant contributing risk factor for pulmonary embolism associated with air travel[46] and Lapostelle considers that the incidence of pulmonary embolism and deep venous thrombosis after long-distance air travel is likely underestimated.[47]

The airport and cabin environment

Airports are an integral aspect of public health protection. Health inspection and sanitation at many airports is the responsibility of local

40 House of Lords Science and Technology Committee on Air Travel and Health, An Update: 1st Report (Session 2007–08) HL Paper 7, 47

41 Danielle Silverman and Mark Gendreau, 'Medical issues associated with commercial flights' (2009) 373 (9680) The Lancet 2067.

42 HM Government, 'Aviation 2050. The future of UK aviation. A consultation' (HM Government Cm 9714 December 2018) 111, para 5.7, citing Civil Aviation Authority, *Airport Accessibility Report 2017/18* (2018).

43 A Qureshi and K M Porter, 'Emergencies in the air' (2005) 22(9) Emergency Medicine Journal 658.

44 Ibid.

45 J A Low and D K Chan, 'Air travel in older people' (2002) 31(1) Age and Ageing 17.

46 Frédéric Lapostolle et al, 'Severe Pulmonary embolism associated with air travel' (2001) 345(11) New England Journal of Medicine 779.

47 Ibid.

public health authorities rather than airport operators.[48] This will inevitably result in very variable local conditions, likely to be poorer in low-resource countries.

The cabin environment itself may represent a public health hazard. The modes of transmission of infectious diseases on board aircraft may be almost identical to those of other indoor environments or enclosed spaces but the aircraft cabin environment facilitates methods of disease transmission. The confined aircraft space, with many common surfaces and limited airflow, provides a favourable environment for infectious disease transmission[49] and airlines are free to set their own rate of air recirculation.[50]

Thornley et al highlight the potential of disease transmission for cabin crew through their work in the cabin, where transmission can recur from the same source over multiple flight sectors: 'infected flight attendants, whether symptomatic or asymptomatic, may have been an ongoing source of contamination of the airplane cabin or of person-to-person transmission to colleagues during their flight sectors'.[51] While other public transportation conveyances will have similar sources, aircraft environments are different given the high surface-to-volume ratios and the relatively small volume-to-passenger ratios.[52]

The limited galley space affects hand-washing practices[53, 54] and the nature of the galley design (compromised space) is affecting safe food-handling practices.[55] Confined spaces inhibit the circulation of workers, which may impair adherence to hygiene standards during food-handling processes and increase the risk of food safety lapses.[56]

---

48 ACI (n 14 above) ch 8, 'Emergency medical services, hygiene and sanitation at airports' 146–148.

49 Hossam Elmaghraby et al, 'Ventilation strategies and air quality management in passenger aircraft cabins: a review of experimental approaches and numerical simulations' (2018) 24(2) Science and Technology for the Built Environment 160.

50 Carol Boyd, *Human Resource Management and Occupational Health and Safety* (Routledge 2004).

51 Craig Thornley et al, 'Recurring norovirus transmission on an airplane' (2011) 53(6) Clinical Infectious Diseases 515.

52 National Research Council, *The Airliner Cabin Environment and the Health of Passengers and Crew* (National Academies Press 2002).

53 Aimee Pragle et al, 'Food workers' perspectives on handwashing behaviors and barriers in the restaurant environment' (2007) 69(10) Journal of Environmental Health 27.

54 Deborah A Clayton and Christopher J Griffith, 'Efficacy of an extended theory of planned behaviour model for predicting caterers' hand hygiene practices' (2008) 18(2) International Journal of Environmental Health Research 83.

55 Ibid.

56 Ilija Djekic et al, 'Food hygiene practices in different food establishments' (2014) 39 Food Control 34.

There is no international, coordinated monitoring body for inflight/onboard hygiene measures. Most airlines set their own cleaning standards although there are minimal regulations through agencies such as the FAA and Occupational Safety and Health Administration (OSHA) in the US.[57] Boyd notes how

> survey findings suggest that airlines have overlooked a number of key areas that are vital to good health and safety practice, and that cabin crew are denied basic rights such as good hygiene, rest breaks and good air quality. Over half of respondents rated hygiene standards on the aircraft as 'poor', and many of their comments blamed short turnaround times, which prevent thorough cleaning of the aircraft.[58]

Food contamination hazards are associated with both food preparation processes on-ground and cabin crew serving meals on aircraft. Foodborne illness issues arise owing to the complexity and confined space, as well as limited sanitary facilities on aircraft.[59]

There are few clear standards for the cleanliness of commercial aircraft cabins. Airlines generally establish their own set of standards, which cleaning companies then follow. Vlagenov detected particularly high viral and bacterial counts on sink faucet handles, worktops or washroom door handles and argued that to minimise the risks for pathogen transmission, cleaning protocols need to be improved and follow strict rules.[60]

With input from industry experts, WHO has produced a *Guide to Hygiene and Sanitation in Aviation* which 'addresses water, food, waste disposal, cleaning and disinfection of facilities, vector control and cargo safety'.[61] This also makes reference to the need for harmonisation with the IHR which requires public health measures at airports and

> the use of scientific principles to prevent, detect, reduce or eliminate the sources of infection and contamination, to improve sanitation in and around international ports, airports and ground crossings, to prevent the international dissemination of vectors and to mandate national and international actions to prevent the international spread of disease.[62]

While these guidelines are helpful, and include sensible recommendations for routine cleaning programmes, training,

57 Scott McCartney, 'The trouble with keeping commercial flights clean' (*Wall Street Journal* 17 September 2014).

58 Boyd (n 50 above).

59 Maija Hatakka, 'Hygienic quality of foods served on aircraft' (Dissertation, University of Helsinki 2000).

60 Kiril Vaglenov, 'Survival and transmission of selected pathogens on airplane cabin surfaces and selection of phages specific for *Campylobacter Jejuni*' (PhD thesis, Auburn University 2014).

61 WHO, *Guide to Hygiene and Sanitation in Aviation* 3rd edn (WHO 2009) 2.

62 Ibid 5–6.

disinfection after an event and the use of protective equipment, they are advisory only. There are no regulations for either the number or size of toilets or washing facilities such as wash basins on an aircraft. Cabin design is not down to aircraft type, but to airline demands and choice. For example, a Boeing 737 is typically configured with three lavatories, but it can also be configured with two or four.[63, 64] Adequate disinfection may be challenging or impossible given the confined space and limited time available. Monitoring of cabin hygiene is not standard practice on many flights. Since the outbreak of COVID-19 many airlines have hastened to announce new and enhanced cabin-cleaning systems[65] and increased passenger seating space. However, this is not an option for budget airlines which follow a business model requiring a high passenger 'load number'. The director general of IATA has argued that social distancing on airlines would mean an end to cheap air travel.[66] There is scepticism that improved hygiene and distancing measures will be adopted long term.[67]

The confined space on an aircraft may also represent a hazard for vulnerable passengers. The average body size of passengers is increasing[68] at the same time that aircraft design has been under pressure to fit in as many passengers as possible to maximise income. Concerns have been raised that this confinement increases the risk for passengers of a deep vein thrombosis (DVT). The WHO WRIGHT project[69] concluded that the risk of developing a venous thromboembolism (which can manifest as a DVT or a pulmonary embolism) doubles after travel lasting four hours or more. Although low, the risk is greater in passengers with predisposing factors such as overweight, use of oral contraceptives, age over 40 years or chronic disease. Vulnerable passengers are estimated to face a two to fourfold risk of DVT on flights of eight hours or longer.[70]

---

63 Collins Aerospace. '737 advanced lavatory'.

64 Dan Reed, 'American airlines' tiny new bathrooms test limits of what US Passengers will put up with' (*Forbes* 30 May 2018).

65 Laura Begley Bloom, 'COVID report: the best and worst airlines during coronavirus' (*Forbes* 27 July 2020).

66 Julia Kollewe, 'Physical distancing will end era of cheap air travel, industry warns' *The Guardian* (London, 21 April 2020).

67 Will Horton, 'New airline seat designs? They won't ever fly on airplanes' (*Forbes* 25 April 2020)

68 Johan Molenbroek, Thomas J Albin and Peter Vink, 'Thirty years of anthropometric changes relevant to the width and depth of transportation seating spaces, present and future' (2017) 65 Applied Ergonomics 130.

69 WHO, *WHO Research into Global Hazards of Travel (WRIGHT) Project: Final Report of Phase 1* (WHO 2007).

70 Roger W Byard, 'Deep venous thrombosis, pulmonary embolism and long-distance flights' (2019) 15(1) Forensic Science, Medicine and Pathology 122.

The risk of DVT is not aviation specific and is almost entirely due to the period of immobility. Those at increased risk are those who have risk factors that apply generally, regardless of the environment. However, no other means of transport are comparable to aircraft with regards to travel time (now up to 18/19 hours), confined seats and restricted movement (particularly those seated in middle rows). It may be a weak risk factor in shorter flights but the risk is likely to increase in long-haul flights.

There are no international regulations concerning the distance between seats (referred to as 'seat pitch') provided to passengers on commercial aircraft, and there may be considerable differences between airlines. Seat pitch has been decreasing since deregulation of the airline industry in the 1970s from around 89 cm to 71–79 cm, depending on the airline and fare class purchased.[71, 72]

There are concerns that insufficient seat pitch will make it difficult for passengers to assume an adequate brace position or evacuate the plane quickly in an emergency.[73] Part of the regulations for certification of any aircraft type/configuration is a requirement for formal testing of the time taken for evacuation of the aircraft and FAA regulations require that commercial aircraft must be evacuated within 90 seconds or less.[74] However, there is criticism that these tests fail to include all members of the population, such as the elderly. Lijmbach et al found that the elderly take significantly more time than younger people during an evacuation.[75] Airlines argue that reductions in seat pitch are necessary to compete with low-cost carriers[76] and Mendoza acknowledged the trade-offs between health risks, price, and airline seat size regulation.[77]

Food contamination

Reports of food contamination are relatively rare, but present a uniquely hazardous event in flight, with potential to incapacitate aircrew as well

71 Scott R Winter, 'Government seat pitch regulation of commercial airlines: a multi-study of consumer perceptions' (2019) 37(2) Collegiate Aviation Review International.

72 Elaine Glusac, 'FAA declines to regulate airplane seat size' *New York Times* (6 July 2018).

73 Claire Quigley et al, 'Anthropometric study to update minimum aircraft seating standards' (Joint Aviation Authorities 2001).

74 Winter (n 71 above).

75 Willem Lijmbach, Peter Miehlke and Peter Vink, 'Aircraft seat in- and egress differences between elderly and young adults' (2014) 58 Proceedings of the Human Factors and Ergonomics Society Annual Meeting.

76 McCartney (n 57 above).

77 Roger Lee Mendoza, 'Health risk, price efficiency, and airline seat size regulation' (2018) 11(2) International Journal of Healthcare Management 122.

as passengers. Most inflight meals are prepared on the ground and then reheated on board. On-ground food hygiene rules are generally strict, governed by national public health laws and in alignment with food preparation regulations in public eating establishments such as restaurants and cafés. Airlines may need to comply with food hygiene regulations from the country where the food is supplied, the country of the airline affiliation, and possibly also public health regulations in the destination country.[78] Apart from the complexity of ensuring compliance, these rules may conflict. Also, while flight catering kitchens are in fact more stringently hygiene-controlled than other on-ground food establishments, facilities to ensure hygiene in food service are limited. The problematic time span is the point where food leaves the catering truck until the aircraft reaches its destination (termed by Sheward as the 'missing link')[79] with little oversight such as audits or compliance controls. Also, airlines rely on local catering companies, with different country standards of food safety.[80] Airline galleys are typically extremely small, and as stated above there are no rules for the size, number or accessibility of hand-washing facilities, such as wash basins.

Cabin crew are classed as professional food handlers,[81] yet research has identified poor aircrew training in food handling[82] and there is little transparency of training programmes for individual airlines.[83] Crew may be interrupted in food preparation by other service demands, for example, if a passenger is unwell and requires attention, yet cabin crew are typically not considered as a vehicle for disease transmission.[84] Food may be left standing or require reheating and maintaining a cold chain may be challenging on ultra-long-haul flights. All of these factors may compromise food hygiene and result in food contamination. There are recorded inflight incidents of food poisoning from agents including salmonella, *Staphylococcus aureus* and *E coli*.[85] However, evidence is limited due to passengers frequently not becoming symptomatic until after arrival in the destination country. Uneaten food is thrown

78 Lauren Solar, 'Food safety takes off: regulations, logistics and the challenges of airline catering' *Global Food Safety Resource*.

79 Erica Sheward, *Aviation Food Safety* (Blackwell 2006).

80 Solar (n 78 above).

81 Ayman Abdelhakim, 'Cabin crew food safety training: an exploratory study' (PhD thesis, Cardiff Metropolitan University 2016).

82 Ayman Abdelhakim et al, 'Cabin crew food safety training: a qualitative study' (2019) 96 Food Control 151.

83 Sheward (n 79 above).

84 Alexandra Mangili and Mark A Gendreau, 'Transmission of infectious diseases during commercial air travel' (2005) 365(9463) The Lancet 989.

85 R McMullan et al, 'Food-poisoning and commercial air travel' (2007) 5(5) Travel Medicine and Infectious Disease 276.

away at the end of a flight and is unlikely to be available for analysis. While there is considerable quality control for on-ground catering kitchens, there is no comparable monitoring of food hygiene inflight. For these reasons, quantifying the incidence of inflight food poisoning is difficult. Furthermore, proving liability is likely to be challenging, except when there has been a mass event where numerous passengers become ill. In economy class the meals are likely to be pre-packaged and pre-prepared (and therefore with less risk of contamination), but this may not be the case in first or business class.

Abdelhakim made an in-depth investigation into cabin crew food safety training and found 'numerous complaints related to food safety and in-flight service ... However, most of these complaints are not available due to the airlines' operations policy.'[86] Long departure delays, length of flight time, and appropriate storage of food at safe temperature zones are all important factors to achieve a true picture of the microbiological quality of food throughout the flight.[87] Incidents of foodborne illnesses among airline passengers are typically investigated in the countries where they occur and by an airline's own quality management team. Health authorities across national borders may neither publish nor monitor foodborne illness rates among passengers.

In the EU, food hygiene is regulated by EC Regulation No 852/2004 although this does not contain any specific reference to aviation. In the UK the Civil Aviation Authority provides a good practice guide.[88] UK-registered aircraft are also required to have a nominated environmental health officer and airlines must also have created a food safety management system.

Outside Europe, national public health regulations will usually be applicable, but there are no harmonised international laws for food safety. The IFSA, in collaboration with WHO, has produced 'World Food Safety Guidelines for Airline Catering', which are based on the Hazard Analysis and Critical Control Point (HACCP) system. HACCP is a science-based system for identifying and responding to specific hazards in food safety. IFSA also plays a role in the audit of flight kitchens (in addition to local authority/national government requirements) on behalf of airlines, to ensure that standards are applied and breaches investigated.

IATA's 'Cabin Operations Safety Best Practices Guide' also provides guidelines on food safety. Both guidelines contain sensible recommendations, but, without robust monitoring, it is difficult to

86 Abdelhakim (n 81 above).

87 McMullan (n 85 above).

88 Civil Aviation Authority, *CAP 757 Occupational Health and Safety On-board Aircraft* (CAA 2012).

ascertain incidence of food contamination, compliance with guidelines or how effective the guidelines have been in reducing risk.

Air quality in the cabin

During the course of a flight the cabin air supply is recycled and filtered regularly using a high efficiency particle air (HEPA) filter, making a complete air change 20 to 30 times per hour. The highest efficiency filters available are comparable to those in hospital operating theatres, catching more than 99 per cent of airborne microbes.[89] However, while airlines generally maintain industry standards and comply with inflight safety regulations, it is not mandated and air quality and circulation rates are susceptible to cost-saving measures in terms of (1) reducing the fresh air provision rate and (2) failing to properly maintain the air-conditioning system.[90]

A currently highly contentious topic is allegations of aircrew becoming ill due to poor air quality in the cabin. These relate to what are known as 'fume incidents', namely, any event in which there is an unusual odour, fume or vapour (other than fire). There are many potential causes and a small proportion may be due to bleed air contamination, that is, the leak of engine oil or hydraulic fluid into the cabin air supply as a result of overfilling or oil seal failure. These fluids contain organophosphates which may become toxic to humans in sufficient quantities. Incidence has been estimated at 0.02 per cent[91] to 0.05 per cent[92] of flights although the seriousness may range from a strong smell to thick smoke.[93] The frequency of these events and causal link with ill effects on passengers and aircrew is strongly disputed.

There have been reports of pilot incapacitation[94] and allegations of ill health resulting in aircrew taking early retirement. In 2010, a flight attendant was successful in the Australian High Court in her claim for

89 IATA, *Briefing Paper: Cabin Air Quality – Risk Of Communicable Diseases Transmission* (IATA Corporate Communications January 2018).

90 Boyd (n 50 above).

91 Maher Shehadi, Byron Jones and Mohammad Hosni, 'Characterization of the frequency and nature of bleed air contamination events in commercial aircraft' (2016) 26(3) Indoor Air 478.

92 Committee on Toxicity of Chemicals in Food Consumer Products and the Environment, 'Statement on the review of the cabin air environment, ill health in air crews and the possible relationship to smoke/fume events in aircraft' (2007).

93 Virginia Harrison and Sarah J Mackenzie Ross, 'An emerging concern: toxic fumes in airplane cabins' (2016) 74 Cortex 297.

94 Sally Evans and Sally-Ann Radcliffe, 'The annual incapacitation rate of commercial pilots' (2012) 83(1) Aviation, Space, and Environmental Medicine 42.

compensation for injury suffered because of contaminated air.[95] Yet, despite many legal claims on behalf of aircrew, proving a causal link is often difficult. Research by EASA in 2017 found that 'cabin/cockpit air quality is similar or better than what is observed in normal indoor environments' and that there was no evidence of a causal link between contaminants and reported ill health.[96] However, Michaelis found that 'a clear cause and effect relationship has been identified linking the symptoms, diagnoses and findings to the occupational environment. Recognition of this new occupational disorder and a clear medical investigation protocol are urgently needed.'[97] The condition was named 'aerotoxic syndrome'. In March 2019 the BBC reported that 51 cases were brought by pilots and cabin crew for ill health arising from exposure to frequent 'fume events'.[98] Defossez argues that if causation can be proven this would 'open the floodgates for litigation' from aircrew[99] who may be repeatedly exposed to contaminated air over their career in the air industry. Passengers would be in an easier position as they could bring a compensation claim for bodily injury from a single incident under article 17 of the Montreal Convention. The aviation industry disputes that air quality is even a public health risk in aviation and 'aerotoxic syndrome' is not recognised in medicine.

The controversy remains ongoing. There is no constant monitoring of cabin air quality so it is difficult to get an accurate estimate of incidence in exposure to contaminants. If the lower statistic of 0.02 per cent incidence is taken, this would mean an incident approximately every 2000 flights, but most studies have used much smaller sample sizes.[100] As reports and legal claims mount there is likely to be increasing pressure for conclusive scientific evidence. If causation can be proven, there will clearly need to be urgent technical innovation to limit this health risk, supported by enforceable regulation. EASA is currently funding further research,[101] and it is to be hoped that the issue can be finally resolved soon.

95 *East West Airlines Ltd v Turner* [2010] HCATrans 238.

96 EASA, 'EASA publishes two studies on cabin air quality' (EASA 23 March 2017).

97 Susan Michaelis et al, 'Aerotoxic syndrome: a new occupational disease?' (2017) 3(02) Public Health Panorama 198.

98 'Airlines face lawsuits over "toxic" cabin air' (*BBC News* 28 March 2019).

99 Delphine Defossez, 'Contaminated air: is the "but for" test saving air carriers?' (2019) 44(2) Air and Space Law 185.

100 Ibid.

101 EASA Tender: EASA.2020.HVP.17: Cabin Air Quality Assessment of Long-Term Effects of Contaminants.

Carriage of disease vectors

Several highly virulent vector-borne diseases have been spread by the carriage of insect vectors on international commercial flights including malaria,[102] West Nile virus[103] and Zika.[104] WHO reported that 'insect vectors may transmit infection to people in places served by aircraft (eg "airport malaria")'. West Nile virus first appeared in the US as a group of cases of patients who lived next to La Guardia airport in New York.[105] The spread of Zika virus to Brazil was attributed to flights from French Polynesia to Brazil during 2013–2014.[106] The problem is being exacerbated by the warmer weather brought by climate change. A major concern is that a new vector may be introduced in an area where it does not currently exist but where the environmental conditions are suitable for the establishment of a breeding population. If aircraft and airports can be kept free of vectors, then the risk of local disease transmission is mitigated.

Annex 5 of the IHR sets specific requirements for vector control:

> 2. Every conveyance leaving a point of entry situated in an area where vector control is recommended should be disinsected and kept free from vectors.

The primary defence against vectors such as mosquitoes is 'disinsection', the use of insecticide sprays inside the aircraft cabin. WHO has produced a list of approved insecticides and guidelines for disinsection procedures,[107] although the use of such insecticides is left to national policy.[108] ICAO similarly leaves the use of insecticides to the discretion of member states, stating only that they should follow WHO recommendations as to the method and procedure to be followed.[109] Disinsection is controversial due to alleged inefficacy and

102 WHO, 'Aircraft disinsection methods and procedures' 25 February 2021.

103 Eleanor B E Brown et al, 'Assessing the risks of West Nile virus-infected mosquitoes from transatlantic aircraft: implications for disease emergence in the United Kingdom' (2012) 12(4) Vector-Borne and Zoonotic Diseases 310.

104 Norman G Gratz, Robert Steffen and William Cocksedge, 'Why aircraft disinsection?' (2000) 78 Bulletin of the World Health Organization 995.

105 Doug Struck, 'Climate change drives disease to new territory' *Washington Post* (5 May 2006).

106 Eduardo Massad et al, 'On the origin and timing of Zika virus introduction in Brazil' (2017) 145(11) Epidemiology and Infection 2303.

107 WHO (n 102 above).

108 Ibid.

109 ICAO, *International Standards and Recommended Practices* 15th edn (October 2017) annex 9 'Facilitation', ch 2, D 2.25.

also potential adverse impacts on aircrew due to constant exposure.[110] Alternative, non-chemical measures have been proposed, such as the use of air curtains, but their effectiveness is still unproven and there is currently no scientific consensus on optimal measures. There is also questionable vector control around airports which is critical to supplement disinsection. Vector control around airports should be implemented both in the airport in the country of departure and in the destination airport where the environmental conditions are sufficiently favourable to enable the establishment of a new population of the vector. Yet vector control at airports is governed (if at all) by local or national regulation. Any measures would need resources and the active cooperation of the departure country, many of which are in low-income settings.

The lack of harmonisation or international consensus has led to some direct conflict of laws. For example, national laws in Australia and New Zealand mandate the use of pyrethroid insecticide on incoming aircraft, while the same insecticide is banned for use in public health in the US,[111] at least partly due to concerns about adverse impacts on the health of aircrew.[112] To our knowledge there has not been any litigation for transmission of insect vectors from an endemic to a non-endemic country, and evidentially this would be difficult to prove against a specific airline, notwithstanding the transmission routes proven by modelling.

Outside environment

Although the present article focuses on public health threats arising within the aircraft cabin, aircraft fuel emissions, noise and waste management are also a public health concern and have been acknowledged as a contributor to climate change.[113] Regulatory bodies have been proactive about this issue. ICAO's annex 16 sets environmental standards to regulate aircraft noise and engine emissions. It established the Committee on Aviation Environmental Protection (CAEP) in 1983 and this now has 'more than 600

110 Binnian Wei, Krishnan R Mohan and Clifford P Weisel, 'Exposure of flight attendants to pyrethroid insecticides on commercial flights: urinary metabolite levels and implications' (2012) 215(4) International Journal of Hygiene and Environmental Health 465.

111 Andrea Grout et al, 'Guidelines, law, and governance: disconnects in the global control of airline-associated infectious diseases' (2017) 17(4) The Lancet Infectious Diseases 118.

112 United States, 'Agenda Item 6: International Health Regulations (IHRs). Non-Pesticidal Disinsection of Aircraft', *ICAO, Facilitation (FAL) Division – Twelfth Session* (Cairo 2004).

113 David S Lee et al, 'Aviation and global climate change in the 21st century' (2009) 43 (22–23) Atmospheric Environment 3520.

internationally recognised experts, in areas such as noise, air quality, climate change but also aircraft end-of-life and recycling and climate change adaptation'.[114] ICAO also developed the Carbon Offsetting and Reduction Scheme for International Aviation (CORSIA) which has a self-imposed target to cut net emissions from aviation fuel by half by 2050 (from its 2005 baseline). The EU has created numerous environmental regulations, including the Environmental Noise Directive,[115] which monitors aircraft noise. Working with ICAO, the EU requires all member states to submit action plans to reduce carbon emissions.[116] IATA has also created a Sustainability and Environment Advisory Council (SEAC),[117] an Environmental Policy[118] and has established programmes to work with airlines to improve their environmental impact.[119]

Although the results are still to be seen, these initiatives represent a responsible and positive response from the aviation industry to international concerns. It would be hugely beneficial if it could take a similar approach to public health threats within the cabin environment itself.

Carriage of infected passengers

While each of the public health risks discussed above merit attention, they are dwarfed by the threat to global health of commercial flights bringing infected persons, whether passengers or aircrew, to non-endemic countries, potentially creating or exacerbating a disease outbreak at local or international level. The threat was realised in February 2003 when the SARS virus was brought by an infected passenger on a flight from Hong Kong to Toronto, Canada, infecting hundreds of individuals, including hospital patients and healthcare workers in that city. The outbreak continued in Canada until June 2003 by which time it had resulted in 438 probable or suspect cases and 43 deaths.[120]

The COVID-19 pandemic has demonstrated the size of the challenge as never before. For Cassar, the growth in aviation traffic made it almost inevitable:

114 ICAO Environment, Environmental Report WHO Aircraft Disinsection.

115 Directive 2002/49/EC of the European Parliament and of the Council of 25 June 2002 relating to the assessment and management of environmental noise.

116 EASA, EEA and EUROCONTROL, 'European Aviation Environmental Report' (2016) 24.

117 IATA Sustainability and Environment Advisory Council.

118 IATA Aviation and Environment Policy.

119 IATA, 'Our actions for the environment'.

120 Bjug Borgundvaag et al, 'SARS outbreak in the Greater Toronto area: the emergency department experience' (2004) 171 Canadian Medical Association Journal.

> Considering the unprecedented volume of travel, specifically by air, and the unprecedented scale of globalization, it comes as no surprise that COVID-19, which has the innate ability of being transmitted easily from one person to another, infected such a large number of people in so many different locations in relatively no time at all.[121]

There are no easy control measures. Passenger screening is of doubtful efficacy, raises difficult ethical questions and is costly in time and resources. Self-reporting is unreliable and passengers may be asymptomatic in the early stages of disease, particularly for diseases with long incubation periods, so are unaware of their infection. It is also difficult for ground crew to spot disease cases and to enforce measures such as denial of boarding. There can be problems if a passenger becomes ill mid-flight due to a lack of space for isolation or quarantine.

Inaccurate pre-departure screening readings, such as temperature, PCR and LFTs, come with substantial consequences at the personal, health system, and societal levels. These include potential virus transmission from an undetected positive case, unjustified cancellation of travel in the case of a false positive result, or even misdirection of policies regarding quarantine and lockdowns.[122] Exit screening may be useful in some instances and was used during the Ebola outbreak of 2014, but will not catch all cases as demonstrated by incidents of aid workers who travelled on commercial flights home to the US and UK and were not diagnosed until after travelling on to their homes. This led to widespread concern, and political questioning of the wisdom of sending aid workers to assist in humanitarian disasters (or at least allowing them to return to their home countries).[123, 124] During the COVID-19 pandemic, passengers on flights from affected areas have been required to enter into 14-day quarantine on arrival and such measures may be of value where there is a known risk, although it remains to be seen what impact this has had on the spread of the virus. It is also still unclear how COVID-19-related quarantine

---

121 Cassar R, 'Evolution or devolution: aviation law and practice after COVID-19' (2020) 45 (Special Issue) Air and Space Law.

122 E Surkova, V Nikolayevskyy and F Drobniewski, 'False-positive COVID-19 results: hidden problems and costs' (2020) 8(12) The Lancet Respiratory Medicine 1167–1168.

123 Stephanie Gee and Morten Skovdal, 'Public discourses of Ebola contagion and courtesy stigma: the real risk to international health workers returning home from the West Africa Ebola outbreak?' (2018) 28(9) Qualitative Health Research 1499.

124 Jenn Selby, 'Donald Trump says Ebola doctors "must suffer the consequences"' *The Independent* (London, 4 August 2014).

and isolation in aviation have collectively affected health equity and human rights.[125]

Airlines and national authorities have the right to refuse passage to persons carrying infectious disease,[126] although in many countries there are also protections against discrimination. For example, in the US it is illegal to refuse to carry a passenger just because they have AIDS.[127] (Although not a public health risk in the context of aviation, HIV/AIDS is still an infectious disease and 'infectious and contagious diseases' are listed under the medical contraindications to flying.) However, unlike AIDS, the greatest public health risks are likely to be from highly contagious diseases spread by droplet or airborne transmission, such as measles and influenza.

IATA guidance recommends that a person onboard who has a suspected communicable disease should be isolated if possible and, if suffering from vomiting and/or diarrhoea, seated near a toilet which should be restricted for use by the ill person(s). However, there may not be adequate space to isolate an ill passenger, especially in the context of higher overall passenger numbers and higher occupancy on each flight. As discussed, sanitary facilities are often limited and will depend on the class and aircraft type, but economy class washrooms are likely to be especially cramped. A passenger who has vomited, perhaps in their seat or in the toilet, may create a public health hazard which cannot be adequately cleaned for several hours, whether because the flight is part-way through a long-haul journey, or because of the pressure for a rapid change over on short flights. Spilt body fluids (blood, vomit etc) must be cleaned up during a turn-round and, if an area cannot be adequately cleaned, for example if fluid has soaked into the fabric of a seat, the area – usually the seat row – should be isolated until such time as this can be adequately dealt with. Many long-haul aircraft will have 'spill kits' for precisely this purpose but the carriage of spill kits is not mandatory and the use of these or any other cleaning practices is neither monitored nor subject to enforceable international regulations.

Pilots are required by annex 9 of the IHR to file a General Declaration at the end of a flight giving notification of any person who has been ill on board and may be suffering from a communicable disease. This will be of limited value if the person has been asymptomatic on board or if the pilot is not fully informed of the person's condition (and therefore

125 WHO 2021, Annexes to Weekly Epidemiological Record (WER). Evidence Review. Public health measures in the aviation sector in the context of COVID 19: Quarantine and Isolation (21 May 2021).

126 Jürgen Graf, Uwe Stüben and Stefan Pump, 'In-flight medical emergencies' [2012] Deutsches Aerzteblatt Online.

127 Federal Aviation Administration, 14CFR Nondiscrimination on the Basis of Disability in Air Travel; Final Rule, 382.21(b)(2).

may not have ensured adequate quarantine or safe transport on arrival for example). A passenger who is infected by a fellow passenger could potentially litigate against an airline (or directly against the fellow passenger), but it could be very challenging evidentially where the ill person is asymptomatic at the point of departure.

As with carriage of vector-borne disease, the lack of scientific consensus over effective control measures undermines and limits any possibility of harmonised regulation.

LIABILITY

Air carriers should owe a duty of care to passengers, aircrew and to populations in destination countries but sometimes these duties may conflict, and current aviation governance is focused on passenger safety. Since the Warsaw Convention of 1929, airlines have owed a legal duty to passengers who have suffered loss, injury or death during an international flight. This is strict but limited liability. This means that, unless there is evidence of contributory negligence, the passenger does not need to prove fault on the part of the airline and therefore takes away the burden of evidential proof. However, damages are limited to 113,100 SDRs (Special Drawing Rights: a currency rate established by the International Monetary Fund).

This legal right was strengthened by article 17 of the Montreal Convention of 1999 which removed the limit on damages, provided an onboard accident causing death or bodily injury can be proven.[128] There has been considerable litigation regarding the meaning of the term 'accident'[129] and it has become defined as 'an unexpected or unusual event or happening that is external to the passenger'.[130] However, airlines can escape liability if they can prove the fault was that of a third party. Along with this duty of care, the Montreal Convention provides that a litigant has a choice of five alternative forums within which to bring a claim.[131]

Aircrew may also bring compensation claims under national legislation for occupational injury suffered during their work. As explained above, compensation has been sought for ill health

128 Montreal Convention 1999, ch III 'Liability of the carrier and extent of compensation for damage'.

129 Ronald I C Bartsch, *International Aviation Law: A Practical Guide* (Routledge 2016) 203.

130 *Air France v Saks* [1985] 470 US 392.

131 Domicile of the carrier; principal place of business of the carrier; country where the contract of carriage was made; destination country; the state of the passenger's principal place of residence (provided the carrier operates, directly or indirectly, to that state).

allegedly sustained through poor cabin air quality and overexposure to insecticides.[132] It might also conceivably be brought for, say, exposure to an infected passenger resulting in crew illness. Aircraft do not carry medical staff as standard practice and aircrew are expected to be 'first responders', perhaps responsible for dealing with a passenger who is vomiting, bleeding or suffers incontinence. Anyone acting in a first-aid role – which is an inherent part of the cabin crew role – is given adequate and appropriate training commensurate with the risk, but there is little insight (if any) into whether or to what extent airline-internal training curricula cover infectious diseases.

Even with the strict, no-fault liability under the Warsaw Convention, for many incidents it will not be possible to show that a passenger or air crew became ill because of a particular journey. Due to long incubation periods, an infectious passenger may be asymptomatic and fellow passengers may not become ill until after arrival and dispersal in the destination country. Although limited damages might be payable under the no-fault terms of the Warsaw Convention, to obtain the more generous compensation under the Montreal Agreement, a passenger would need to prove that this was due to the fault of the airline. Evidential difficulties might arise in trying to prove that a particular illness was contracted due to the condition of the cabin interior, say, due to dirty toilets, food trays or tray tables.[133]

However rare an event, carriage of disease vectors has brought highly dangerous diseases across the world. The risk might be reduced by effective disinsection and improved environmental controls at airports. Yet, despite WHO guidelines, current measures for disinsection are often haphazardly executed and of doubtful efficacy[134] and environmental control of vectors at airports is challenging in many high-risk endemic countries due to limited public health capacity. There is unlikely to be successful legal action against either airlines for allowing vectors on board (due to the virtual impossibility of proving that a particular vector was carried on a particular aircraft) or against national authorities for the same difficult evidential reasons.

DISCUSSION

While the right to legal recourse for injury is important, the priority should be on prevention and response to public health threats. The aviation industry has the advantage of an ethos and structures for strong international and multi-agency collaboration. This has already

132 'Qantas steward with Parkinson's to sue over pesticide link' *Bangkok Post* (9 December 2013).

133 Vaglenov (n 60 above).

134 Grout et al (n 111 above).

shown itself to be responsive to new environmental threats and a similar approach should now be taken to other public health threats in aviation. 'Harmonisation' should not be confused with 'uniformity' of laws since it allows for national sovereignty in interpretation, but there need to be baseline, enforceable common standards. The benefits of harmonised laws may seem self-evident but are worth repeating. International aviation, by definition, crosses national borders, so public health threats are cross-border health threats affecting several countries. Harmonisation of laws means that the rules to deal with these threats will be the same across all jurisdictions, thereby reducing administrative complexity and avoiding conflict of laws where countries may have conflicting rules, for example on food handling or the use of insecticides. It ensures consistency and allows for greater collaboration between member states and industry stakeholders in the creating of these laws. It should be an opportunity to create high standards which follow the best scientific evidence and respect human rights. The latter aspect is important because some of these regulations will have ethical aspects – for example on isolating passengers, or refusing to carry them on aircraft. Above all, aviation is a global industry like none other for its potential to damage global health, and this demands the highest possible international standards.

So how might a robust, effective and harmonised regulatory regime be created to improve public health protection in aviation? Is it even possible to achieve the same hygiene standards as are enforced on the ground? Commercial pressures and lack of national capacity are likely to be obstacles, but those have been successfully overcome in the past in order to achieve a high level of mechanical safety.

Without scientific consensus it will be very difficult to require countries to introduce new measures and comply with harmonised standards. The greatest threats to public health – carriage of vectors infected with human pathogens or infectious passengers – are also the most challenging to resolve. Nevertheless, COVID-19 may provide the impetus and research investment that is needed. There are also less intractable problems which might be addressed more quickly and easily. Poor hygiene and poor sanitation are clearly public health issues. Onboard cleaning largely depends on airline-internal protocol and most airlines set their own cleaning standards. There is only minimal regulation through agencies such as the FAA and OSHA[135] and no monitoring bodies.

Regulations for toilets and perhaps improved, ergonomic design to make cleaning easier should be possible. The cost of providing adequate space for isolation of an ill passenger might be a difficult

135 McCartney (n 57 above).

barrier, given the low number of incidents but better aircrew training should be possible. Following the example of hospitals, hand sanitisers might be provided throughout the cabin for the use of passengers as well as crew. This is justifiable in light of the unique form of transport, confined space conditions, and the fact that people from all over the world share this space.

There needs to be better identification of the types of food most at risk of contamination, examination of behaviours for food preparation and service, monitoring and enforceable regulations to bring standards into line with those on the ground.

The SARS outbreak led to closer cooperation between WHO, ICAO, IATA and subsequently ACI. SARS was also a catalyst to the 2005 revision of the IHR which incorporated numerous references to aviation. Similarly, ICAO updated its SARPS to recommend greater preparedness at airports and the need for member states to develop national health plans to deal with public health emergencies.[136] It also led directly to ICAO establishing the Collaborative Arrangement for the Prevention and Management of Public Health Events in Civil Aviation (CAPSCA), a collaboration of regulatory bodies to review the spread of communicable diseases which has declared that:

> Coordinating the international aviation response to public health risks, such as pandemics, is a key role for the International Civil Aviation Organization. By means of international, regional, national and local organizations are brought together to combine efforts and develop a coordinated approach.[137]

While encouraging and valuable, these remain matters of guidance only, and are not prescriptive.

Cuinn and Switzer argue that public health emergencies such as SARS and Ebola have led to a more coherent governance framework, pointing to the joint development of a Passenger Locator Form in 2007 and the Traveller Public Health Declaration Form during the Ebola outbreak, both self-reporting forms which are used to improve surveillance and tracing of potentially infected passengers. However, they accept that there remained 'something of a "gap" when it came to governing a crucial component of the aviation sector – the interior of the cabin and infected passengers'.[138] They also found that:

> the aircraft is a site of legal contestation. Tensions were revealed between the intersections of legal systems. These were particularly prevalent when it came to the collection and handling of passenger data and

136 Cuinn and Switzer (n 1 above)

137 CAPSCA Collaborative Arrangement for the Prevention and Management of Public Health Events in Civil Aviation.

138 Cuinn and Switzer (n 1 above).

> were only partly resolved by the bridging work performed between the regimes. This reveals a legal plurality within the constitutive assemblage of global health security; a finding which has significant implications for the development of international responses to infectious disease.[139]

It may be argued that ICAO's remit is primarily passenger safety and that public health is a matter for public health authorities – and at international level this means the WHO. It would certainly need a major realignment of responsibilities for aviation regulators to take the lead on this, going beyond their current role of collaboration with public health authorities on public health risks linked to aviation. Yet the evidence that aviation has facilitated the spread of successive global outbreaks, leading to the global catastrophe that is COVID-19, surely now demands a debate on responsibilities.

An alternative regulatory actor, at least at regional level, might be the EU. EASA's 'European Plan for Aviation Safety (EPAS) 2019–2023' has a goal to 'achieve constant safety improvement within a growing aviation industry'. This policy document includes proposals to address environmental factors such as aircraft emissions and aircraft noise. EASA's research on cabin air quality shows that it is willing to investigate cabin safety issues, even if that particular public health threat remains unresolved. With political will, adequate funding and legislative authority over 27 EU member states and four European Free Trade Area states,[140] EASA has the potential to make a significant impact on public health protection in aviation. Either ICAO or EASA might expand the mandates of their environmental bodies to include the cabin environment.

CONCLUSION

Due to a scarcity of available data, the present article can only provide an outline of the legal framework for aviation, examples of some key regulations and a snapshot of a few identified public health risks. Until COVID-19, the aviation industry was growing exponentially, with vulnerability increasing in tandem and disease incidents becoming more frequent – SARS, Zika, and now COVID-19. Existing governance structures require a radical rethink and overhaul to ensure they can adequately manage these vulnerabilities.

Above all else, there is a need for a comprehensive and reliable quantification of risks. This would require far more data than are currently collected, including detailed monitoring of the cabin

139 Ibid.

140 Switzerland, Norway, Iceland, Leichtenstein. As a result of Brexit, the UK is no longer a member of EASA.

environment and passenger demographics. There will be cost and feasibility implications but, at minimum, they should include better access to passenger data, including greater access to incident and accident data. Such information would enable more reliable risk analyses, allowing problems to be prioritised and informing approaches to tackling them. A full systematic review of the legal landscape and mapping of responsibilities is also required along with collaboration to agree responsibilities at international level.

For decades the aviation industry has been a model to others for its collaborative approach to regulation, enabling aviation to become one of the safest forms of transport. It now needs to draw on that strength to tackle the enormous challenges ahead of it. The devastating impact of the COVID-19 pandemic calls for a fundamental reassessment of the roles of the WHO and ICAO in managing public health risks in aviation.

Northern Ireland
Legal Quarterly
Autumn Vol. 74.3 (2023) 619–642
Commentaries and Notes DOI: 10.53386/nilq.v74i3.1084

The Safe Access Zones Bill Reference [2022] UKSC 32: clearing the fog of proportionality?

Anurag Deb*

Queen's University Belfast
Correspondence email: adeb01@qub.ac.uk

INTRODUCTION

In March 2022, the Northern Ireland Assembly passed the Abortion Services (Safe Access Zones) Bill (Northern Ireland) (SAZ Bill)[1] to create buffer zones around lawful abortion providers, in an attempt to criminalise the harassment and intimidation of people who seek services offered by such places or work in them. This is the first such legislative measure anywhere in the United Kingdom (UK) or Ireland, with Scotland and Ireland exploring equivalent measures.[2]

In the interim, the Attorney General for Northern Ireland (AGNI) referred the SAZ Bill to the UK Supreme Court to determine whether it was consistent with the rights set out in the European Convention on Human Rights (ECHR), and thus within the Assembly's legislative competence.[3] On 7 December 2022, the UK Supreme Court handed down judgment in the *Reference by the Attorney General for Northern Ireland – Abortion Services (Safe Access Zones) (Northern Ireland) Bill* (*SAZ Reference*).[4]

The two major issues for the Court were the appropriate approach to proportionality and to *ab ante* challenges to legislation. The first issue required consideration of the Court's previous judgment in *Ziegler*[5] and

* A much-condensed version of the section on appellate proportionality in this article appears in Anurag Deb, 'The Safe Access Zones Bill Reference: appellate proportionality' (2023) 27(2) Edinburgh Law Review 225–232. I am grateful to Dr David Capper for his thoughts on an initial draft. I am also grateful to Matthew McCallion BL and the staff at the Bar Library, Belfast, for indulging me in my search for a niche but completely anodyne point of Northern Ireland legal history.

1 Royal Assent granted on 6 February 2023. As the judgment concerns the (then) Bill, this article refers to the Bill and not the Act.

2 For Scotland, see Scottish Government, 'Abortion Safe Access Zones: Ministerial Working Group'. For Ireland see the Safe Access to Termination of Pregnancy Services Bill 2021.

3 Northern Ireland Act 1998, s 6(2)(c).

4 [2022] UKSC 32, [2023] 2 WLR 33.

5 *DPP v Ziegler and Others* [2021] UKSC 23, [2022] AC 408.

the judgment of the Divisional Court (England & Wales) in *Cuciurean*.[6] The second issue required consideration of two additional precedents: *Christian Institute*[7] and *Re McLaughlin*.[8] Unusually for a devolution reference, the Supreme Court sat as a panel of seven justices. The *SAZ Reference* judgment was unanimous and delivered by Lord Reed.

The length of this article reflects both the length and complexity of the judgment. The issues explored by the Court are not only myriad, but each issue is also underpinned by multiple decisions of the highest domestic authority. These decisions at times appear to overlap and at other times appear to contradict one another. The *SAZ Reference* attempts to tie these loose ends into a single coherent approach. In what follows, I attempt to explore whether the judgment succeeds in that endeavour.

THE BILL PROVISIONS

The SAZ Bill has four main interrelated components.

First, it defines 'protected premises' which are healthcare facilities[9] where information, advice or counselling in relation to abortion services are provided[10] and the operator of such a facility has notified the Northern Ireland Department of Health of the intention for the facility to be protected as such.[11]

Second, the SAZ Bill defines 'protected persons' as anyone attending protected premises to access treatment, information, advice or counselling,[12] anyone accompanying a person seeking such access[13] or anyone working at such premises.[14]

Third, the Bill establishes 'safe access zones', defined as the public area within at least 100 metres[15] (extendable to 150 metres)[16] from each entrance to and exit from protected premises.

Fourth, and the main part for the Court, the Bill creates two offences within safe access zones: the first is the criminalisation of any act with intent or recklessness as to whether that act influences a protected person, prevents or impedes their access to protected

6 *DPP v Cuciurean* [2022] EWHC 736 (Admin) [2022] QB 888.
7 *The Christian Institute v Lord Advocate* [2016] UKSC 51, 2017 SC (UKSC) 29.
8 [2018] UKSC 48, [2019] NI 66.
9 SAZ Bill cl 2(2).
10 Ibid cl 2(3).
11 Ibid cl 2(4).
12 Ibid cl 3(a).
13 Ibid cl 3(b).
14 Ibid cl 3(c).
15 Ibid cl 4(2).
16 Ibid cl 4(3).

premises or causes them alarm, harassment or distress.[17] The second offence criminalises the recording of a protected person without their consent within a protected zone, with intent or recklessness as to whether that recording has any of the same effects as the first offence.[18] Both offences are summary offences only, punishable with a fine of up to £500.[19]

THE PRELIMINARY POINT: STATUTORY INTERPRETATION AND LEGISLATIVE COMPETENCE

The Court considered the test which the Bill must pass in order to be within the Assembly's competence. There is an important point to be made here, considering that this was a devolution reference[20] and not a post-enactment challenge grounded on a specific and real factual matrix. This is thus an *ab ante* challenge, in respect of which the Supreme Court had, in *Christian Institute*, asked whether the legislation under challenge 'is capable of being operated in a manner which is compatible with [the ECHR] rights in that it will not give rise to an unjustified interference … in all or almost all cases'.[21]

In the *SAZ Reference*, the Court pointed to a tension between the test in *Christian Institute* and a later citation of it in *Re McLaughlin*. *McLaughlin* was a challenge to the provision of widowed parent's allowance being paid to surviving spouses but not surviving *unmarried* partners under article 8 of the ECHR (read with article 14).[22] Here, the *Christian Institute* test was cited by Lady Hale, who referred to legislation operating incompatibly in 'a legally significant number of cases'.[23] In the *SAZ Reference*, Lord Reed indicated that this was an inaccurate citation of the *Christian Institute* test and reiterated its original form as accurate.[24]

With respect, this appears to be a problematic reading of the relevant passages across the two cases. In *Christian Institute*, the Court had been concerned with the requirement that legislation should operate *compatibly* in all or almost all cases, leaving open the possibility that compatible legislation may nevertheless operate incompatibly in some cases. In *McLaughlin*, by contrast, the reference to 'legally significant'

17 Ibid cl 5(2).
18 Ibid cl 5(3).
19 Ibid cl 5(4).
20 Northern Ireland Act 1998, s 11(1).
21 *Christian Institute* (n 7 above) [88], the court citing *R (Bibi) v Home Secretary* [2015] UKSC 68, [2016] 2 All ER 193.
22 *McLaughlin* (n 8 above) [1].
23 Ibid [43].
24 *SAZ Reference* (n 4 above) [19].

was to cases where legislation may operate *incompatibly*.[25] Thus, contrary to how the AGNI had characterised *McLaughlin* as being 'less demanding' than *Christian Institute*,[26] *McLaughlin* was instead the corollary to *Christian Institute*: if legislation operated incompatibly in a legally significant number of cases, it cannot be said to operate compatibly in all or almost all cases, thus failing the test in *Christian Institute*. Read in this way, the Court's reiteration of the *Christian Institute* test in the *SAZ Reference* seems unnecessary, especially as regards the 'clarification' of Lady Hale's words in *McLaughlin*.[27]

CLARIFYING *ZIEGLER* AND *CUCIUREAN*

At issue for the Supreme Court was whether the criminalisation of influencing a protected person[28] disproportionately interfered with three ECHR provisions: article 9 and the protection of religious freedoms, article 10 and the protection of free speech and expression and article 11 and the protection of free assembly. Central to this question was the issue of proportionality. The Court, therefore, began not with the Bill, but with *Ziegler* and *Cuciurean*. There were two main points underlying the Court's consideration of both cases: proportionality was a legal test and not a factual one, and that general legal prohibitions (such as might be enacted in statutes) may be proportionate in themselves without requiring a proportionality analysis on a case-by-case basis. There is a great deal of detail and complexity in the discussion of both cases, including the historical approaches to 'lawful excuse' or 'reasonable excuse' defences. At the heart of this complexity, however, is a simple question: does an offence interfering with free speech need a 'reasonable excuse' defence in order to be a proportionate interference?

The Court's consideration of *Ziegler* begins with a pointed observation: that the remarks of Lords Hamblen and Stephens (who delivered the majority judgment in *Ziegler*) about proportionality being a 'fact-specific enquiry … requir[ing] the evaluation of the circumstances in the individual case' should not be considered a universal rule. Instead, these remarks should be confined to the trial of offences under section 137 of the Highways Act 1980 (wilful

25 *McLaughlin* (n 8 above) [43]: 'the test is not that the legislation must [original emphasis] operate incompatibly in all or even nearly all cases. It is enough that it will inevitably operate *incompatibly* [emphasis added] in a legally significant number of cases.'

26 *SAZ Reference* (n 4 above) [12].

27 The court might also have considered that Lady Hale's remarks in *McLaughlin* were an almost exact reproduction of her remarks in *Bibi* (n 21 above) [60], which itself was the origin of the test in *Christian Institute* (n 7 above) [88].

28 SAZ Bill, cl 5(2)(a).

obstruction of a highway without lawful authority or excuse), where ECHR rights under articles 9, 10 and 11 were raised.[29] The same point was made by the Divisional Court in *Cuciurean* (more on that further below).

This is, with respect, a strange observation. *Ziegler* in the Supreme Court was concerned with answering two questions certified for appeal by the Divisional Court, the first of which asked what the proper appellate approach was to offences containing a 'lawful excuse' defence when engaging ECHR rights.[30] While the second question was concerned with the section 137 offence, it followed the first question, in that the first question asked for a *general* test, and the second question asked for that test to be *specifically* applied. In the *SAZ Reference*, this point seems to have eluded the Court's criticism of one of the intervenors' (JUSTICE) position that *Ziegler* was (at the very least) capable of being read as having laid down a universal rule.[31]

Substantively, the first question in *Ziegler* asked the Supreme Court about the appropriate way in which proportionality should be judicially assessed – whether as a question of fact, the answer being appealable only for an error of law (favoured by the majority) or a question of law which should be appealable in any event (favoured by the minority). In the *SAZ Reference*, it was the minority view in *Ziegler* which prevailed, but with the additional point that general legislative measures may themselves be proportionate without being evaluated against the specific factual circumstances of a particular case.[32]

But here, the Court in the *SAZ Reference* was faced with the European Court of Human Rights' decision in *Perinçek v Switzerland*,[33] in which the Grand Chamber specifically stated that in interferences with free speech which lead to criminal convictions,

> ... it is normally not sufficient that the interference was imposed because its subject-matter fell within a particular category or was caught by a legal rule formulated in general terms; what is rather required is that it was necessary in the specific circumstances.[34]

In the *SAZ Reference*, the Court focused on the word 'normally'. *Perinçek* was a case concerning the criminalisation in Swiss law of genocide denial as a disproportionate interference with article 10 rights. The Grand Chamber had pointed to the Swiss Government's acceptance that criminalisation needed to be balanced against free speech and expression in individual cases 'in such a way that only

29 *SAZ Reference* (n 4 above) [28]–[29].
30 *Ziegler* (n 5 above) [7].
31 *SAZ Reference* (n 4 above) [28].
32 Ibid [34].
33 (2016) 63 EHRR 6.
34 Ibid [275].

truly blameworthy cases would result in penalties' and that the Swiss courts had not 'paid any particular heed to this balance'.[35] In the *SAZ Reference*, the Court used these passages ostensibly in order to distinguish *Perinçek* from the SAZ Bill.[36] However, as will become clear further below, there are certain circumstances surrounding the passage of the SAZ Bill which muddy its distinction from *Perinçek*.

Two further points followed. First, that the European Court does not 'review legal provisions and practice in abstracto' but confines itself to scrutinising the application of the ECHR in the case before it, whereas the Supreme Court could not proceed on this basis in 'a reference of the present kind' – a reference to the *ab ante* challenge (to which I return further below).[37] Second, that in order to give the ECHR rights a practical and effective dimension, the Court could not make a distinction in the application of the ECHR to civil and criminal measures, by reference to the Government's practice (post-*Ziegler*) of obtaining 'persons unknown' injunctions in respect of protestors rather than prosecuting them under relevant statutory offences.[38]

The second point is ostensibly a reference to *Perinçek*, but it is somewhat problematic. The Court rejected the idea that it should take a particular approach to proportionality in a criminal context. It did so by pointing to 'persons unknown' injunctions as civil remedies, which they are – a civil remedy with potentially *criminal consequences* if breached. Seen in this light, the point which the Court made – that there should be no difference in approaching proportionality between civil and criminal measures – disappears if the focus turns from the nature of the measure to the consequence of breaching it. *Perinçek*, importantly, concerned the Grand Chamber because of the severity of the consequences for breaching the Swiss law in question.[39]

Cuciurean received different treatment from *Ziegler*; given that the Divisional Court's position on *Ziegler* in *Cuciurean* aligned with that of the Supreme Court in the *SAZ Reference* (as set out earlier), there was no real need to clarify the impact of the latter case. The Supreme Court did, however, lay down general guidance on how to approach proportionality issues in criminal trials where rights under articles 9, 10 and 11 of the ECHR are raised. First, there is a question whether those rights are engaged at all, considering certain acts (for example, incitement to violence or criminal damage to property) fall outside the scope of those rights.[40] Second, the question arises whether the

35 Ibid [275]–[276].
36 *SAZ Reference* (n 4 above) [39].
37 Ibid [40].
38 Ibid [41].
39 *Perinçek* (n 33 above) [272].
40 *SAZ Reference* (n 4 above) [54].

ingredients of the offence(s) themselves satisfy the proportionality requirement.[41] Third, if the ingredients of the offence do not satisfy proportionality, then the trial court may use the interpretive duty under section 3 of the Human Rights Act 1998 to make the offence ECHR-compliant or assess the proportionality of a conviction if the offence is statutory.[42] If the offence is a common law offence, the court may 'develop the common law so as to render the offence compatible with Convention rights'.[43]

A case which the Supreme Court did not consider in its judgment, but which was (at least) referred to in oral argument was *Lee Brown v PPSNI*.[44] *Brown* was an appeal by way of case stated in the Northern Ireland Court of Appeal, concerning the proportionality of a conviction for publishing or distributing threatening, abusive or insulting written material[45] against the defendant's article 10 rights.[46] The defendant had been convicted of the offence, with the District Judge being satisfied that the conviction was proportionate. An appeal to the County Court was dismissed. The Court of Appeal allowed the case stated appeal on the basis that the District Judge had not considered or balanced the competing interests between the prosecution and the defendant's ECHR rights.[47] The absence of *Brown* is odd when considering that it was handed down by a member of the *SAZ Reference* panel – the Lady Chief Justice of Northern Ireland. More substantively, however, *Brown* was important for two reasons. First, it applied *Ziegler* in a way which the Divisional Court had held to be incorrect in *Cuciurean*.[48] Thus, there was an obvious tension between high judicial authority in different UK jurisdictions. Second, the Court of Appeal in *Brown* had favoured the approach to the role of appellate courts in *Ziegler* over its own broader statutory jurisdiction to decide questions of fact for itself.[49] The role of appellate courts following *Ziegler* was not considered in *Cuciurean* – only whether *Ziegler* had or had not laid down a universal rule.

The net effect of the Court's consideration of *Ziegler* and *Cuciurean* was therefore twofold. First, that case-by-case proportionality analyses are unnecessary where a defendant raises issues under articles 9–11 of the ECHR. Second, where a proportionality analysis is carried out, it is

41 Ibid [55].
42 Ibid [57].
43 Ibid [61].
44 [2022] NICA 5.
45 Public Order (Northern Ireland) Order 1987, art 10(1).
46 *Brown* (n 44 above) [1]–[2].
47 Ibid [75]–[77].
48 Ibid [63]. See *DPP v Cuciurean* (n 6 above) [67].
49 Ibid [65].

not a purely fact-dependent exercise. But to what extent does this mark a change in the legal understanding of proportionality?

A DEEPER DIVE INTO PROPORTIONALITY JURISPRUDENCE

It is worth exploring the Court's scrutiny of *Ziegler* to see what (if anything) needed a critique or clarification in that case. As will become clear further below, this scrutiny bore significant consequences for the Court's assessment of the SAZ Bill.

There are three interrelated issues in the Court's scrutiny of *Ziegler*: the nature of a proportionality assessment under the ECHR, the role of appellate courts when faced with proportionality assessments and the use of precedent in *Ziegler* itself.

As to the nature of a proportionality assessment, Lord Reed began with the position that proportionality 'is not an exercise in fact-finding', citing Lord Bingham's remarks in *A v Home Secretary* in support.[50] Lord Bingham, for his part, stated that 'the European Court does not approach questions of proportionality as questions of pure fact'.[51] While this is uncontroversial, it is worth recalling what happened in *A*. Lord Bingham criticised the Court of Appeal's approach to the Special Immigration Appeals Commission's (SIAC) proportionality assessment as being 'unappealable findings of fact' and allowed the appeal on the basis that SIAC's reasoning based on its findings of fact was vitiated by errors of law.[52] The House relied on SIAC's findings of fact concerning a threat to national security;[53] it was SIAC's reasoning as to whether those findings *justified* the discriminatory measures in issue (and the Court of Appeal's endorsement of this reasoning) which the House of Lords overruled.

The role of appellate courts was a central aspect of the majority's reasoning in *Ziegler*, and which the Court in the *SAZ Reference* clarified. In the latter, the Court favoured a more interventionist approach by appellate courts when faced with questions of proportionality than the approach purportedly adopted in *Ziegler*.[54] But to what extent was this evaluation of *Ziegler* accurate? This is not a straightforward or simple point, but it is important to explore it in some detail.

---

50 *A v Home Secretary* [2004] UKHL 56, [2005] 2 AC 68, cited in *SAZ Reference* (n 4 above) [30].

51 Ibid [44].

52 As Lord Bingham stated (ibid): 'The reasons given by SIAC do not warrant its conclusion … I do not consider SIAC's conclusion as one to which it could properly come.' .

53 Ibid [27].

54 *SAZ Reference* (n 4 above) [33].

To begin, the Court in the *SAZ Reference* took issue with *Ziegler*'s analysis of the role of an appellate court for two main reasons. First, the lack of reference in *Ziegler* to those cases which the Court in the *SAZ Reference* identified as demonstrating a more 'interventionist' approach: *Baiai*,[55] *Nicklinson*,[56] *UNISON*,[57] *SC*,[58] *A*, *Bank Mellat (No 2)*[59] and *Elan-Cane*.[60] Second, the reliance in *Ziegler* on a *dictum* of Lord Carnwath in *R*,[61] which the Court in the *SAZ Reference* said was context-specific (to that case), to the effect that an appellate court should not interfere in the proportionality assessment conducted by a lower court merely because the appellate court would have arrived at a different evaluation. I take each point in turn.

On the first point, the comparison the Court drew in the *SAZ Reference* between the seven 'interventionist' cases and the approach favoured by the majority in *Ziegler* is less clear than at first glance. Six out of the seven cases all either identified legal errors which vitiated the proportionality assessments of lower courts,[62] or agreed that the proportionality assessments by lower courts were legally sound.[63] Appeals were allowed in the former category and dismissed in the latter. The decision in *Nicklinson* was unusually complex, both factually and legally. On the issue of proportionality, the Supreme Court was concerned that the courts below had been deprived of the evidence and argument needed for a full assessment of proportionality,[64] and that at least some of these matters were first presented before the Supreme Court itself. Consequently, it is difficult to say with any certainty whether the Supreme Court's approach in *Nicklinson* was more interventionist than that in *Ziegler*; in a major way, it had acted as the court of first instance when fully assessing proportionality in *Nicklinson*. The key point here, however, is that in the remaining six

55 *R (Baiai and Ors) v Home Secretary* [2008] UKHL 53, [2009] 1 AC 287.

56 *R (Nicklinson and Another) v Ministry of Justice; R (AM) v DPP; R (AM) v DPP* [2014] UKSC 38, [2015] 1 AC 657.

57 *R (UNISON) v Lord Chancellor* [2017] UKSC 51, [2020] AC 869.

58 *R (SC, CB and Others) v Work and Pensions Secretary* [2021] UKSC 26, [2022] AC 223.

59 *Bank Mellat v Her Majesty's Treasury (No 2)* [2013] UKSC 39, [2014] 1 AC 700.

60 *R (Elan-Cane) v Home Secretary* [2021] UKSC 56, [2022] 2 WLR 133.

61 The reference in the SAZ Reference judgment contains the neutral citation [2018] UKSC 47, which is the case of *R(AR) v Greater Manchester Police* [2018] UKSC 47, [2018] 1 WLR 4079. The paragraph references in the *SAZ Reference*, as well as the précis of the case facts given by Lord Reed in that judgment all match those in *AR*. If this is a typographical error, the error may originate in the report of the case in the Weekly Law Reports. The rest of this article refers to *AR* instead of *R*.

62 *UNISON* (n 57 above) [112], *A* (n 50 above) [44] and *Bank Mellat (No 2)* (n 59 above) [27].

63 *Baiai* (n 55 above) [28], *SC* (n 58 above) [71] and *Elan-Cane* (n 60 above) [62].

64 *Nicklinson* (n 56 above) [120].

cases, the House of Lords or the Supreme Court only interfered with the proportionality assessments of lower courts where there was a legal error which vitiated those assessments. Turning to the approach in *Ziegler*, Lords Hamblen and Stephens said:

> ... an appeal will be allowed where there is an error of law material to the decision reached which is apparent on the face of the case, or if the decision is one which no reasonable court, properly instructed as to the relevant law, could have reached on the facts found. In accordance with that test ... where the statutory defence depends upon an assessment of proportionality, an appeal will lie if there is an error or flaw in the reasoning on the face of the case which undermines the cogency of the conclusion on proportionality.[65]

If there is *per se* a distinction between the approach in *Ziegler* and that in the six cases cited by the Court in the *SAZ Reference*, it is far from clear. This is especially true of *Brown*, which followed *Ziegler*: it is difficult to see how a more interventionist approach to proportionality would have changed the outcome.

On the second point, the Court in the *SAZ Reference* warned against attaching 'undue significance to a statement which was made by Lord Carnwath (in *AR*) in the context of a particular case without reference to a plethora of other cases'.[66] These 'other cases' were references to the seven cases explored above. In *AR*, Lord Carnwath said:

> The decision [of the lower court] may be wrong, not because of some specific error of principle in that narrow sense, but because of an identifiable flaw in the judge's reasoning, such as a gap in logic, a lack of consistency, or a failure to take account of some material factor, which undermines the cogency of the conclusion. However, it is equally clear that, for the decision to be 'wrong' ... it is not enough that the appellate court might have arrived at a different evaluation.[67]

It is important to understand these remarks in context: they conclude a section of Lord Carnwath's judgment entitled 'proportionality in the appellate court'[68] in which he examined multiple prior authorities on this point. It is unnecessary to delve into all these authorities, but they all share a common strand, which Lord Carnwath adopted as his conclusion above. Indeed, one of these cases, *In re B*,[69] is particularly germane to the discussion here. *In re B* was concerned with care orders under the Children Act 1989, but its discussion of the proper appellate approach to proportionality foreshadowed the same discussion in *Ziegler* with an uncanny resemblance. The majority

65 *Ziegler* (n 5 above) [54].
66 *SAZ Reference* (n 4 above) [33].
67 *R(AR) v Greater Manchester Police* [2018] UKSC 47, [2018] 1 WLR 4079, [64].
68 Ibid [53].
69 [2013] UKSC 33, [2013] 1 WLR 1911.

approach on this issue (Lords Wilson, Neuberger and Clarke) aligned with the majority in *Ziegler*[70] and the minority approach (Lord Kerr and Lady Hale) aligned with the minority in that case.[71] Now, in the *SAZ Reference*, the Court restricted the impact of *In re B* by pointing to the case being about specific care orders.[72] But the fact of the case concerning care order proceedings operated differently. Lady Hale considered that the paramountcy of the welfare of children under the Children Act 1989 was, together with the duty under section 6(1) of the Human Rights Act 1998, a strong reason to favour an appellate court deciding proportionality for itself.[73] Lord Kerr also tied his reasons to a combination of the section 6(1) duty with the specific context of proceedings involving children.[74] The majority's approach, however, was concerned with a general approach to proportionality in an appellate setting. Thus, the fact of *In re B* being a care order case mattered for the minority rather than the majority – contrary to how it was evaluated in the *SAZ Reference*.

Moreover, Lord Carnwath did not simply cite *In re B* as dispositive of the question. He buttressed his view with the general function of an appellate court as explored by Lord Reed in *McGraddie v McGraddie*, to the effect that an appeal is an opportunity to correct lower court errors rather than reargue a case.[75] Thus, far from a decision which is context-specific, *AR* drew multiple proportionality analyses into an attempt to provide a general approach.

This discussion of proportionality jurisprudence takes us to a critical case discussed in *Ziegler*: *Edwards v Bairstow*.[76] *Edwards* concerned a tax assessment in connection with the sale of a Yorkshire spinning plant, raising the question whether the first instance tax assessment was a matter with which appellate courts could (and should) interfere. The High Court and Court of Appeal both determined that the assessment was untouchable except if legally perverse.[77] The House of Lords strongly disagreed and allowed the appeal. A passage in Lord Radcliffe's speech on the proper approach to appeals was cited in *Ziegler*,[78] but it is worth setting out a passage in Viscount Simonds' speech about the nature of inferences derived from the facts of a given case.

70 See Ibid [46] per Lord Wilson JSC, [88] per Lord Neuberger PSC and [136] per Lord Clarke JSC.
71 See ibid [118] per Lord Kerr and [205] per Lady Hale JJSC.
72 *SAZ Reference* (n 4 above) [33].
73 *B* (n 69 above) [204].
74 Ibid [121].
75 [2013] UKSC 58, 2013 SLT 1212, [3], cited by Lord Carnwath JSC in *AR* (n 67 above) [57].
76 [1956] AC 14.
77 Ibid 19–20.
78 *Ziegler* (n 5 above) [37], citing *Edwards* (n 76 above) 36.

> ... it must be clear that to say that such an inference is one of fact postulates that the character of that which is inferred is a matter of fact. To say that a transaction is or is not an adventure in the nature of trade is to say that it has or has not the characteristics which distinguish such an adventure. But it is a question of law not of fact what are those characteristics, or, in other words, what the statutory language means. It follows that the inference can only be regarded as an inference of fact if it is assumed that the tribunal which makes it is rightly directed in law what the characteristics are ...[79]

Applying these remarks to *A*, we see that the threat to UK national security emanating from terrorism was a fact, but this did not justify implementing measures to combat terrorism *only* against foreign nationals. Here, the question whether a fact justifies a measure (or, to use Viscount Simonds' language, whether a fact is justificatory in character) is a question of law and thus subject to appellate scrutiny. This is precisely what the House of Lords did in *A*.[80] It is difficult therefore to see why the approach in *Edwards*, as endorsed by *Ziegler*, was differentiated in the *SAZ Reference* at all, far less differentiated as being less interventionist than cases such as *A*.[81]

However, this is not to say that the Court's differentiation in the *SAZ Reference* was completely without foundation. Lord Sales, in the minority in *Ziegler*, differentiated between *Edwards* and proportionality assessments by stating, 'the legal standard being applied in the former is the standard of rationality and in the latter is the standard of proportionality'.[82] It seems clear that the Court in the *SAZ Reference* had a similar view, at one point referring to the *Ziegler* approach as being 'a standard of unreasonableness when considering issues of proportionality'.[83] It is worth setting out Lord Sales' own reflections on what unreasonableness or rationality meant in a *judicial* context:

> ... the difference between application of the ordinary rationality standard on an appeal to identify an error of law by a lower court or tribunal and the application of the proportionality standard for that purpose in a context like the present should not be exaggerated.[84]

There is an important reason why. Assuming amenability to judicial review, the application of the 'ordinary' rationality standard only allows

79 *Edwards* (n 76 above) 30–31.

80 *A* (n 50 above) [44] per Lord Bingham.

81 *SAZ Reference* (n 4 above) [33].

82 *Ziegler* (n 5 above) [137].

83 *SAZ Reference* (n 4 above) [33].

84 *Ziegler* (n 5 above) [138], referring to Lord Carnwath's remarks in *AR* (n 67 above) [64]. The *Ziegler* majority did not distinguish rationality and proportionality and instead pointed to a line of authorities exploring the nature of criminal appeals as grounded partly in *Wednesbury* rationality, see *Ziegler* (n 5 above) [29]–[35].

a court to interfere in discretionary non-judicial decisions where the decision in question is robbed of logic.[85] This is because a court does not self-evidently possess the capacity and expertise for decision-making in any context other than a judicial one; a judge is qualified in law and not policy.[86] This is not the same as an appellate court interfering in the decision of a lower court, because both possess the same capabilities over legal reasoning. Thus, the *Ziegler* approach entails a level of scrutiny which is, by its very nature, more 'interventionist' than the High Court reviewing the decision of a minister or a local authority. Although Lord Sales warned in *Ziegler* against treating rationality and proportionality interchangeably,[87] it is clear that the lines between them are blurred.[88]

A related issue is whether a straight line can be drawn between *Edwards*, rationality and *Ziegler*, as the Court appears to have done in the *SAZ Reference*. In *Ziegler*, Lords Hamblen and Stephens explored the 'conventional' approach of the Divisional Court to appeals by way of case stated (of which *Ziegler* was one) as involving rationality, citing a number of Divisional Court judgments in support.[89] But none of these judgments cited *Edwards*. Lords Hamblen and Stephens themselves did not equate *Edwards* and rationality outright, merely observing that *Edwards* is an authority for appellate restraint in connection with findings of fact, and that appellate restraint is also exhibited by the Divisional Court judgments.[90] Lord Sales in *Ziegler* cited Lord Diplock's equation of *Edwards* with rationality in *Council for Civil Service Unions v Minister for the Civil Service*,[91] but even this remark was *obiter* in that case, as Lord Diplock had invoked *Edwards* to justify the court's interference with irrational decisions, rather than using it to define the rationality standard itself.

---

85 *R v Parliamentary Commissioner for Administration, ex p Balchin* [1996] 1 PLR 1 (EWHC) [27], per Sedley J (as he then was).

86 *Council of Civil Service Unions v Minister for the Civil Service* [1985] AC 374, 411F, per Lord Diplock.

87 *Ziegler* (n 5 above) [138].

88 See eg *Kennedy v Charity Commission* [2014] UKSC 20, [2015] 1 AC 455, [55]–[56] per Lord Mance JSC, endorsed in *Pham v Home Secretary* [2015] UKSC 19, [2015] 1 WLR 1591, [60] per Lord Carnwath JSC and [109] per Lord Sumption JSC. For academic commentary cited with approval in these judgments, see P Craig, 'The nature of reasonableness review' (2013) 66(1) Current Legal Problems 131 and G Lübbe-Wolff, 'The principle of proportionality in the case-law of the German Federal Constitutional Court' (2014) 34 Human Rights Law Journal 12.

89 *Ziegler* (n 5 above) [29]-[35].

90 Ibid [37]–[39].

91 *Council of Civil Service Unions* (n 86 above) 410H–411A. Cited in *Ziegler* (n 5 above) [137].

I now turn to the *SAZ* Court's final issue with *Ziegler*: the use (or non-use) of precedent. Lords Hamblen and Stephens in *Ziegler* had also been critical of the Divisional Court (in the same case) for not referring to a number of authorities which they considered relevant.[92] On one level, one Supreme Court panel identified a set of authorities which it considered relevant to a given issue, while another panel of the same court identified another set of authorities which that panel considered relevant to the same issue. Logically, the larger panel prevailed.

However, whether or not a judicial panel missed relevant authorities is not the issue here. No panel can conceivably examine *every* authority on a point of law before deciding it and requiring such an exercise would strain resources, reason and possibly even sanity. Rather, the issue is one of framing. The *SAZ Reference* and *Ziegler* framed proportionality in palpably different ways. In the former, Lord Reed's anchor for proportionality lay in the court's 'constitutional function and … its duty under the Human Rights Act'.[93] In the latter, the majority was instead focused on the *appellate* approach to proportionality assessments (which, after all, was one of the two questions certified for appeal in *Ziegler*). *AR*, another decision which had its effect restricted by the *SAZ Reference*, also explored the issue of appellate review of proportionality assessments.

However, just because the difference in the framing of proportionality between the two cases is palpable, it does not follow that the difference is consequential. The crux of the criticism of *Ziegler* is contained in a short section of the *SAZ Reference*, in which the Court galloped through around 15 years of proportionality jurisprudence in various factual contexts. By exploring this jurisprudence on a granular level, however, it is apparent that there is not much clear blue water between the two approaches. The duty not to act incompatibly with ECHR rights under the Human Rights Act does not, by itself, turn an appeal from a review to a rehearing, and not even the *SAZ Reference* suggests otherwise. Thus, the real difference between the two cases is one of degree. If, as explored earlier, these different approaches effectively produce if not the same then similar results (in terms of when appellate courts interfere with lower courts' proportionality assessments), then how significant is the degree of difference between them?

Of course, one may point to the difference in means between *Ziegler* and the *SAZ Reference*, rather than the result. *Ziegler* asked appellate courts to review lower courts' proportionality assessments while the *SAZ Reference* directed appellate courts to conduct proportionality assessments themselves. But of the interventionist precedents cited in

92 *Ziegler* (n 5 above) [29]–[35].

93 *SAZ Reference* (n 4 above) [33].

the *SAZ Reference*, three[94] reviewed the lower courts' proportionality assessments for errors while three[95] approached proportionality largely or completely independently of the lower courts.[96] Thus, even in the means, there is no bright line distinction of the kind drawn in the *SAZ Reference* between its favoured approach and that of *Ziegler*.

A paper by Lord Sales sheds further light on the existence and extent of this distinction.[97] In the paper, Lord Sales expanded on the argument he subsequently made in *Ziegler*, observing that there is a difference in appellate approaches between deference to first instance proportionality assessments and appellate courts conducting such assessments themselves.[98] It is unnecessary to embark on a critical evaluation of the paper in extensive detail for present purposes. What is relevant in the present context is the reason why Lord Sales saw a problem with the approach exemplified in cases such as *AR* and *Ziegler* and what his proposed approach is.

In Lord Sales' view, the tension between the duty to act compatibly with the ECHR and the general rule in England and Wales that an appeal is by way of review rather than rehearing, was resolved *In re B*:

> The obligation of the appellate court under section 6 of the HRA [Human Rights Act] did not require it to depart from its normal appellate function under CPR [Civil Procedure Rules] part 52.11, of secondary review of the trial judge's decision.[99]

A major problem with adopting this as a general approach to proportionality in an appellate setting, according to Lord Sales, is that the nature of appeals varies in the UK's different jurisdictions. As proportionality is itself a general rule, it must rise above the nature of an appeal according to a distinct approach found in England and Wales only.[100]

Instead, Lord Sales proposed the following approach:

> The appellate court should adopt a primary decision-making function when it is able to add value to the normative exercise in deciding whether a measure can be regarded as proportionate, where that potential for added value sufficiently reflects the additional costs and delay associated with an appeal.

94 *Bank Mellat (No 2)* (n 59 above) [22], *A* (n 50 above), [44], and *Baiai* (n 55 above), [27]–[28].

95 *Elan-Cane* (n 60 above) [56]–[61], *SC* (n 58 above) [56] –[60], and *UNISON* (n 57 above) [90]–[99].

96 *Nicklinson* (n 56 above), as set out earlier, stands out uniquely in this group of interventionist cases.

97 Lord Sales, 'Proportionality review in appellate courts' (2021) 26(1) Judicial Review 40.

98 Ibid 40–42.

99 Ibid 49.

100 Ibid 50.

> Generally, that is unlikely to be the case in relation to reviewing facts found by the first instance court. But the appellate court has a constitutional function to articulate and police general legal norms. Thus, there may be a spectrum of potential engagement by an appellate court, depending on the precise nature of the issue which arises in relation to a proportionality assessment. On this approach, there will be differences of degree, regarding how far the appellate court should be drawn into acting as primary decision-maker to make the assessment afresh for itself. Depending on the circumstances in a particular case, it may be possible for the appellate court to accept findings of fact made at first instance (subject only to rationality review) and then supply its own view of the values in contest in that factual position and of the normative outcome.[101]

If we revert briefly to Viscount Simonds' speech in *Edwards*, there is little difference between these two approaches. In *Edwards*, the finding of fact was not in issue – it was whether the fact could give rise to the impugned inference in that case; in short, whether the Income Tax Commissioners' view on the *legal consequence* of that fact could be upheld. The House in that case supplied its own view of the legal consequence of the finding of fact, thereby overruling the Commissioners. In doing so, it also adopted a primary decision-making function as the authoritative expositor of law.

Edwards also provides an answer to Lord Sales' concerns about hewing too closely to the CPR when the nature of an appeal may differ across the UK. The case was decided at a time when appeals were conducted by way of rehearing in England and Wales[102] and Northern Ireland,[103] the latter having retained this approach to this day.[104] The adoption of the *Edwards* approach into the modern appellate setting, therefore, does not necessarily privilege the CPR – certainly, any reference to the CPR would be wholly unjustified in a Northern Ireland appeal to the Supreme Court. Rather, the approach in *Edwards* takes account of appellate restraint in a system where appellate courts have an important public role in maintaining confidence in the legal system itself.[105] This restraint, and its importance in the public role of an appellate court, operates regardless of jurisdictional differences in the general nature of an appeal. This explains why the Northern Ireland Court of Appeal declined to utilise its statutory jurisdiction to its fullest extent in *Brown*, observing that to do so would be 'inappropriate in

101 Ibid 57.

102 Rules of the Supreme Court 1883, order LVIII r 1.

103 Rules of the Supreme Court (Northern Ireland) 1936, order LVIII r 1.

104 Rules of the Court of Judicature (Northern Ireland) 1980, order 59, r 3(1).

105 *Edwards* (n 76 above) 38 per Lord Radcliffe. See also J A Jolowicz, 'Appeal and review in comparative law: similarities, differences and purposes' (1986) 15(4) Melbourne University Law Review 618, 631.

an appeal of this nature',[106] speaking to the legal system in general, rather than the interests of the particular appellant.

The above discussion bears some similarities with the perceived divergence between the English and Scottish courts' approaches to the question in issue in *Edwards*. Lord Radcliffe demonstrated that there was no real divergence at all, given that the relevant authorities in both jurisdictions largely aligned on the issue: the legal consequence of a finding of fact was self-evidently a matter for appellate scrutiny and interference if that consequence was wrongly determined by the courts below. The real divergence, if any, was 'in the understanding and application of the governing principles'.[107] To that end, Lord Radcliffe pithily summed up the proper appellate approach:

> Their duty is no more than to examine those facts with a decent respect for the tribunal appealed from and if they think that the only reasonable conclusion on the facts found is inconsistent with the determination come to, to say so without more ado.[108]

When considered together with the discussion of reasonableness or rationality in a judicial context (as above), the extent to which this approach is inappropriate for proportionality is, at best, debatable.

We thus arrive at the end of a meandering journey through two decades of proportionality jurisprudence to find that appellate courts have arguably been aligning rather than diverging in their approaches. In the circumstances, and with the greatest respect to the Supreme Court, if the *SAZ Reference* appears to have clarified little, perhaps there was little to clarify in the first place.

THE PROPORTIONALITY OF THE SAZ BILL

The SAZ Bill was moved in the Assembly to remedy a serious situation. Vulnerable, anxious women and those who assisted them in accessing abortion services, advice and counselling, were spat at, assaulted, verbally abused and splashed with holy water. Clare Bailey, the former leader of the Northern Ireland Green Party who introduced the SAZ Bill in the previous Assembly, described her own experience at the receiving end of 'a very deliberate campaign of harassment and intimidation against women'.[109]

Having been introduced and voted through its second stage, the SAZ Bill came before the Assembly Health Committee for consideration.

106 *Brown* (n 44 above), [65] per Keegan LCJ.
107 *Edwards* (n 76 above), 38.
108 Ibid 39.
109 *SAZ Reference* (n 4 above) [91].

The Committee Report[110] makes for interesting reading, with one of the most relevant aspects being a remark from then Health Minister Robin Swann MLA. The Bill as introduced had conferred discretionary powers on the Minister's Department regarding the location and extent of a safe access zone. The Minister made it clear that he did not think such discretionary powers were appropriate for the Department. The full passage is worth setting out:

> The Minister advised that in making these decisions, his Department would become responsible for balancing the safety and dignity of protected persons and the right to respect for private and family life on the one hand against the right to manifest religious belief and the rights to freedom of assembly and expression on the other. In the Minister's view, these are not appropriate functions for the Department of Health, as it does not, and should not, have competence in this arena and stating that such matters are therefore *better left to the judicial system*. (emphasis added)[111]

Matters being left to judicial discretion, including the operation of a 'reasonable excuse' defence to clause 5(2)(a), were developed further in the fourth stage of the Bill's passage through the Assembly, where an amendment which would have added a 'reasonable excuse' defence was supported by the Bill's sponsor and defeated by four votes.[112]

These points demonstrate that the debate surrounding the proportionality of the SAZ Bill's offences in general, and clause5(2)(a) in particular, were complex, nuanced and decided, on the point of a proposed defence, on a knife-edge. This bears similarities with the deliberations surrounding the Swiss law in *Perinçek*, which the Supreme Court distinguished in the *SAZ Reference*. These nuances and complexities, moreover, were accounted for in the positions of the Lord Advocate and JUSTICE in respect of the question of the Assembly's competence over clause 5(2)(a). Both parties invited the Court to declare that the clause *was* within competence, *inter alia* because a conviction under this clause would nevertheless be subject to the trial court's obligations under the Human Rights Act, and thus enabled a proportionality analysis of any conviction on a case-by-case basis.[113]

Turning to the Court's consideration of the main issue of the Bill's proportionality, its analysis was concise and uncomplicated on the majority of the relevant questions. The Court held, rather

110 Northern Ireland Assembly, Report on the Abortion Services (Safe Access Zones) Bill (NIA 133/17-22) (27 January 2022) para 86.

111 Ibid [86].

112 Northern Ireland Assembly, Official Report, 49 (Amendment 4) (14 March 2022).

113 *SAZ Reference* (n 4 above) [6] and [9] for the positions of the Lord Advocate and JUSTICE, respectively.

straightforwardly (and unsurprisingly), that clause 5 of the SAZ Bill restricted rights under articles 9–11 of the ECHR,[114] that these restrictions were prescribed by law (ie the Bill itself)[115] and that the Bill pursued a legitimate aim – that of ensuring access to abortion facilities for treatment, advice and employment,[116] and further that of ensuring access to healthcare.[117] On several of these issues, the parties were also agreed.

In its assessment of whether the Bill's restrictions were necessary in a democratic society, the parties agreed (and the Court, with them), that the Bill's aim (ensuring access for protected persons) was sufficiently important to justify interferences under articles 9–11.[118] Moreover, the Court held that there was a rational connection between the Bill's aim and the means by which it sought to achieve that aim.[119]

The Court's deliberations on the third and fourth proportionality questions, however, were more elaborate. The third question (whether there were less restrictive alternative means available than those in the Bill) was answered affirmatively, with the Court noting that the Assembly had debated and rejected the 'reasonable excuse' defence, *inter alia*, because of the possibility of the defence being used (and abused) to effectively nullify the Bill's aim.[120]

The fourth question (whether the Bill struck a fair balance between individual rights and the interests of the community) received the most detailed answer. Rather than setting out each factor the Court considered relevant to answering this question, these factors were divided into three broad categories. First, the impact of protest, influence and behaviour which might satisfy the requirements of harassment on women seeking to access abortion services or advice on those services, or employees of those services (which, it is important to remember, provide those services *lawfully*).[121] Second, the restrictions imposed by the Bill on rights under articles 9–11 were themselves spatially limited: the offences under clause 5 were not outright bans throughout Northern Ireland[122] and the penalties were monetary (and limited) rather than custodial.[123] Third, the ECHR grants a wide margin of appreciation in

114 Ibid [111]–[112].
115 Ibid [113].
116 Ibid [114].
117 Ibid [115].
118 Ibid [117].
119 Ibid [118].
120 Ibid [121]–[123].
121 Ibid [125]–[126] and [128].
122 Ibid [127].
123 Ibid [130].

matters involving 'sensitive and controversial questions of ethical and social policy' such as abortion.[124]

These points provided answers to the AGNI's concerns about bans on individual protests surrounding abortion as stifling public debate on the issue as well as the criticism of the extent of the safe access zones as defined in the Bill.[125] Both concerns were rejected by the Court by pointing to the spatially limited nature of the offences under clause 5, observing that the 100- to 150-metre limits were not unjustified.[126]

Relatedly, in oral argument,[127] counsel for JUSTICE (Blinne Ní Ghrálaigh) provided some examples of factual circumstances which may warrant a proportionality analysis on a case-by-case basis. The first was that a safe access zone within 150 metres of an abortion clinic may unavoidably extend to sites unrelated to abortion clinics.[128] Even a 100-metre safe access zone would exclude numerous other businesses and an extension of such a zone may even exclude notable sites of public gathering such as Belfast City Hall. Another example was of a silent protest within a safe access zone at such an early hour that most (if not all) staff would not even be present. Both circumstances could technically engage behaviour prohibited by the clause 5 offences, but to what extent would their enforcement be a proportionate interference with ECHR rights? The point of this is not to argue against the aim of the Bill, but to set out what a trial court may be faced with in a prosecution under clause 5.

But these matters were not explored by the Court in its judgment. Instead, the Court drew on judgments in similar matters across a range of jurisdictions, including British Columbia, Ontario, Victoria, Tasmania and a Dutch case determined by the erstwhile European Commission on Human Rights.[129] Here, the Court's reasoning deserves a deeper dive. The Court's framing of justification ('proportionality *stricto sensu*')[130] was to ask whether the clause 5 offence was a fair balance between the rights of access to abortion services and the right to protest against the provision of these services.[131] To that end, all of the comparative jurisprudence on which the Court drew involved conducting balancing exercises in light of *specific* facts – the constitutionality of the criminal laws at issue

124 Ibid [131].
125 Ibid [132]–[134].
126 Ibid [133].
127 See Day 2 (20 July 2022), Afternoon Session of the SAZ Reference hearing.
128 This is true of at least one such clinic (in a busy high street) which I pass by daily on my morning commute to work.
129 *SAZ Reference* (n 4 above) [141]–[153].
130 Lord Sales (n 97 above), 42.
131 *SAZ Reference* (n 4 above) [124].

was determined with respect to how these laws operated in a factual context specific to each case. The *SAZ Reference*, by contrast, could only be concerned with how the SAZ Bill *would* operate, rather than drawing from real practice. This is of course the nature of an *ab ante* assessment of the proportionality of a bill. But this is also why the assessment of proportionality in such circumstances requires a degree of circumspection. After all, we are concerned here with a general measure (strict liability offences) and the practical operation of such a measure is material to the assessment of its proportionality.[132]

An illustrative example relevant to the *SAZ Reference* is the practice of seeking 'persons unknown' injunctions, with their breaches being dealt with by way of contempt proceedings.[133] The courts' jurisdiction to punish for contempt is general in the sense that any injunction granted must be obeyed without exception,[134] but any decision to punish, as well as the punishment itself, must be proportionate having regard to the factual circumstances of the breach, including those of the alleged contemnor.[135] A finding of contempt for breaching a single injunction, therefore, may be proportionate in one factual situation but not another.

This is why the Court's finding that the clause 5 offences are *inherently* proportionate[136] has the effect of shutting down even the possibility of hard cases to test these provisions where they matter most: practice. This marks a considerable departure from the genesis of the *ab ante* test in *Bibi*, which concerned an immigration rule requiring pre-entry English competence on the part of foreign spouses of British citizens (or those settled in the UK).[137] The Supreme Court drew a distinction between the rule operating disproportionately in specific cases and the rule being inherently disproportionate, recognising that the former did not necessarily result in the latter,[138] all without foreclosing the possibility of cases of specific (future) disproportionate operation. The question in *Bibi* was whether a law was capable of operating proportionately. But it does not follow that, just because a law is capable of operating proportionately in all or almost all cases, it is *incapable* of operating disproportionately in specific cases. This is where specific operational examples from practical legal operation

132 *Animal Defenders International v United Kingdom* (2013) 57 EHRR 21, [108].
133 *SAZ Reference* (n 4 above), [41].
134 See eg *Cuciurean v Transport Secretary and Another* [2021] EWCA Civ 357, [9(4)], per Warby LJ.
135 Ibid [17]. See also eg *MBR Acres and Others v McGivern* [2022] EWHC 2072 (QB), [96] per Nicklin J.
136 *SAZ Reference* (n 4 above), [155].
137 *Bibi* (n 21 above), [1].
138 Ibid [2] per Lady Hale DPSC and [69] per Lord Hodge JSC.

assume importance, with the Court in *Bibi* squarely acknowledging the requirement of 'examination on the facts of specific cases' in order to determine whether a law declared *ab ante* proportionate may still operate disproportionately.[139] This aside, it is important to recall that, in *Bibi*, the Court considered examples of how the law in question may operate disproportionately before determining whether the law was *ab ante* proportionate.[140] The Court in the *SAZ Reference*, by contrast, drew together *general* principles from cases decided by courts in other jurisdictions. This is not the same thing as considering any specific circumstances in which the SAZ Bill would operate (especially when, as previously set out, the Court was invited to consider some of these circumstances). It is therefore especially jarring that the Court should have concluded that the SAZ Bill's offences were inherently proportionate.

This is not to argue that the SAZ Bill was disproportionate in the circumstances of the reference. Rather, the point is that proportionality is not predictable. Although the Court was at pains to point out the democratic credentials of the Bill and the margin of appreciation accorded by the European Court to matters such as abortion access,[141] such credentials must necessarily be caveated. A validly enacted law may operate in a changed legal landscape[142] and, as set out earlier, the debates accompanying the Bill's passage were not clearly decisive as far as the clause 5 offences were concerned. In these circumstances, it is worth remembering that the mere passage of legislation does not preclude a judicial assessment of its compatibility with ECHR rights.[143] At the risk of being accused of judicial supremacism, neither legislative arithmetic nor the quality of legislative debates ensures a law operates proportionately in all cases, the more so because the process of law-making is not the same as the process of legal interpretation.[144]

The language of clause 5 certainly achieved an appropriate balance between protest and access to healthcare in the examples of threatening, abusive, intimidating and violent behaviour presented to the Court. But this is very different from saying the clause 5 offences could *never* operate disproportionately, especially considering that the law has yet to come into force. To hold that no proportionality assessment of any prosecution of these offences is required is to omit this reality.

This omission, however, is understandable as a consequence of the Court's earlier minimisation of the importance of fact-finding

139 Ibid [73] per Lord Hodge JSC.
140 Ibid [50]–[55] per Lady Hale DPSC.
141 *SAZ Reference* (n 4 above) [131] and [140].
142 *Wilson v Trade and Industry Secretary* [2003] UKHL 40, [2004] 1 AC 816.
143 *A* (n 50 above) [42].
144 *SC* (n 58 above) [169].

in proportionality. If proportionality is not fact-specific, then the assessment of a general measure in specific operational examples is unnecessary, which is the Court's ultimate conclusion.[145] But proportionality is fact-dependent. This dependency is writ large in one of the main doctrines underpinning the jurisprudence of the European Court: the margin of appreciation. Here, the European Court defers to national authorities' evaluation of the 'local needs and conditions' in which the ECHR must be given effect.[146] Thus, for example, despite an emerging European consensus favouring broader access to abortion services than were available in Ireland,[147] the European Court nevertheless paid particular attention to the 'profound moral views of the Irish people' in dismissing a claim that Ireland's (then) highly restrictive abortion provisions breached article 8 of the ECHR.[148] The Irish law on abortion was thus proportionate not generally, but on the particular facts of Irish society at the relevant time.[149]

The SAZ Bill is compatible with the ECHR not necessarily because it is inherently proportionate, but because it would operate proportionately in almost all cases. Foreclosing the possibility of disproportionate operation (however rare or infrequent) marked not only a departure from the established approach to *ab ante* challenges, but also omitted the fact that a case-by-case proportionality analysis is rooted in a statutory duty – section 6 of the Human Rights Act. This omission is curious given the prominence of the Human Rights Act in the appellate approach to proportionality favoured by the Court (as discussed earlier).

CONCLUSION

In law, the framing of a question is critical. The Court's concluding remarks on its judgment in the *SAZ Reference* provide some insights on its framing of the substantive question of the SAZ Bill's ECHR compliance:

> The right of women in Northern Ireland to access abortion services has now been established in law through the processes of democracy. That legal right should not be obstructed or impaired by the accommodation of claims by opponents of the legislation based, some might think ironically, on the liberal values protected by the Convention. A legal system which enabled those who had lost the political debate to

145 *SAZ Reference* (n 4 above) [155].

146 *Buckley v United Kingdom* (1996) 23 EHRR 101, [74].

147 *A, B and C v Ireland* (2011) 53 EHRR 13, [235].

148 Ibid [241].

149 This has remained the view of the European Court, see eg *RR v Poland* (2011) 53 EHRR 31, [187].

> undermine the legislation permitting abortion, by relying on freedom of conscience, freedom of expression and freedom of assembly, would in practice align the law with the values of the opponents of reform and deprive women of the protection of rights which have been legislatively enacted.

The Court thus framed the question as a balancing exercise between access to reproductive healthcare and the expression of opinions on the availability of that healthcare. Although the Court commendably recognised the serious situations faced by vulnerable women accessing reproductive healthcare, its remarks are odd in the context of this case for two reasons. First, the *SAZ Reference* was a devolution reference validly taken by a relevant law officer in circumstances where none of the parties or intervenors sought to 'undermine' legislation permitting abortion[150] (the SAZ Bill, of course, does not permit abortion services but merely protects access to them). Second, the Court's framing of proportionality and its approach to the *ab ante* challenge to clause 5 led to an ironic outcome. The judgment simultaneously asked for greater judicial intervention in criminal matters while precluding all such intervention into the SAZ Bill's own criminal provisions.

In the end, the *SAZ Reference* was a significant milestone in the history of women's rights in Northern Ireland. Access to safe and lawful abortion services is a matter of reproductive healthcare and Northern Ireland's history in this respect is viewed by many as a textbook case in gender discrimination.[151] More widely, however, its legacy might lie in its application to increasing legislative trends towards criminalising protests.[152] As the enacted law hardens in this context, judicial scrutiny appears to have commensurately softened.

150 Bearing in mind the Supreme Court's recent endorsement of a devolved law officer's ability to bring references before it (even on a prospective bill) in *Reference by the Lord Advocate of devolution issues under paragraph 34 of Schedule 6 to the Scotland Act 1998* [2022] UKSC 31, [2022] 1 WLR 5435.

151 Committee on the Elimination of Discrimination against Women, *Report on the Inquiry concerning the UK under article 8 of the Optional Protocol to the CEDAW* (6 March 2018) [64]–[83].

152 Public Order HL Bill (2022–23).

Northern Ireland
Legal Quarterly
Autumn Vol. 74.3 (2023) 643–656
Commentaries and Notes DOI: 10.53386/nilq.v74i3.1101

Liability in nuisance: *Fearn v Board of Trustees of the Tate Gallery* [2023] UKSC 4

Francis McManus

University of Stirling
Correspondence email: francis.mcmanus@stir.ac.uk

ABSTRACT

The author discusses the recent Supreme Court case of *Fearn v Tate Modern Galleries*. In *Fearn*, the court was required to determine whether the defendants' allowing visitors to the viewing gallery, which was situated at the top of the Tate Modern, to stare into domestic flats, which were situated close to the Tate, constituted a nuisance in law. The claimants' flats were of an unusual design, in that the external walls which faced the Tate, were constructed entirely of glass, thereby allowing visitors to the Tate to stare into the interior of the flats. By a bare majority, the court held that such a use of the defendants' premises ranked as a nuisance. Whereas the majority of the court upheld the traditional view that, for a claimant to succeed in a nuisance action, the use of the defendant's land required to be unreasonable, in order to determine whether that use was unreasonable, one was required to ascertain whether the defendant's use of land had caused substantial interference with the ordinary use of the claimants' land. However, in turn, the claimant could not complain if the use which was interfered with was not an ordinary use of that land. The court held that the use of the viewing gallery had caused a substantial interference with the ordinary use and enjoyment of the claimants' property.

The majority of the Supreme Court concluded that both lower courts had erred by laying store by the fact that the use of the viewing gallery was of public benefit. However, public interest was a factor which required to be addressed only when the court was ascertaining whether to grant an injunction or an award of damages.

The author argues that the dissenting judgment of the court is to be preferred over that of the majority, most importantly for the following reasons. It may be difficult first, to determine whether the defendant's use of land deviates from the norm, and therefore does not rank as a 'common and ordinary' use of land, and secondly it may prove difficult to weigh such use of land against that of the claimant. The author argues that the test of reasonableness as a test for liability in nuisance, as hitherto employed by the courts, is more conducive to clarity. The traditional test also allows the law to develop both coherently and incrementally, by considering the changing norms of society.

Keywords: nuisance; intrusive viewing; interference with common and ordinary use of land; viewing galleries; public interest; remedies.

BACKGROUND AND DECISION

At the top of the Tate Modern Gallery (the Tate) there was a public viewing gallery, which was opened to the public in 2016. Unfortunately, visitors to the viewing gallery could see straight into the living areas of the claimants' flats, which were situated in close proximity (about 34 metres) to the Tate. The walls of the flats which faced the Tate were constructed mainly of glass. The claimants sought an injunction, requiring the defendants, namely, the Board of Trustees of the Tate, to prevent members of the public from viewing their flats from the relevant part of the Tate or, alternately, an award of damages. The claim was based on the private law of nuisance.

The trial judge found that the intrusive viewing from a neighbouring property could give rise to a claim for nuisance. However, he held that the intrusion which the claimants experienced, did not amount to a nuisance, on the basis that the Tate's use of the viewing gallery was reasonable. The trial judge also found that the claimants were responsible for their own misfortune, firstly, because they had bought property with glass walls, and secondly because they could have taken remedial measures to protect their own privacy, such as lowering their blinds during the day, or installing net curtains.

The claimants appealed. The Court of Appeal dismissed the appeal. The claimants appealed to the Supreme Court which (by a majority of three to two) allowed the appeal. As the trial judge and the Court of Appeal both found no liability in nuisance, the remedy question, injunction or damages did not need to be addressed. Liability was the only issue in the Supreme Court, so the allowing of the appeal required the remedy question to be remitted to the trial court.

In the Supreme Court, Lord Leggatt (who gave the majority opinion) stated that the tort of private nuisance protected a claimant not from the physical invasion of the claimant's land itself, but rather, from the resulting interference with the utility, or amenity value, of the claimant's land.[1] Moreover, there was no requirement that the interference was caused by a physical invasion of the land. He went on to state that there was no reason why a state of affairs, which consisted of the defendant allowing his premises to be used as a base for members of the public to stare into neighbouring property, could not be actionable as a private nuisance.[2]

Lord Leggatt stated that, whereas it was sometimes said that, as a governing principle, to give rise to liability in nuisance any interference with the claimant's enjoyment of land had to be unreasonable, 'unreasonableness' was not itself a legal standard or test, which assisted

1 [2023] UKSC 4, [13].
2 Ibid [17].

one in concluding that a nuisance existed. There were principles, settled since the nineteenth century, which govern whether use of the claimant's land was unreasonable.

In applying these principles, the first question which one was required to ask was whether the defendant's use of land had caused a *substantial* interference with the *ordinary* use of the claimants' land.[3] Lord Leggatt stated that the test for 'substantial' was objective.[4] Furthermore, what amounted to material or substantial interference was to be judged by the standards of an ordinary or average person in the claimants' position. Lord Leggatt went on to state that the objective nature of the test reflected the fact that the interest protected by the law of private nuisance was the utility of land and not the bodily security or comfort of the particular individuals occupying it.

Lord Leggatt stated that an occupier could not complain if the use which was interfered with was not an ordinary use of land.[5] The other aspect of the core principle was that, even where the defendant's activity substantially interfered with the ordinary use and enjoyment of the claimants' land, the activity would not give rise to liability if the activity itself was no more than the ordinary use of the defendant's own land.[6]

Lord Leggatt then stated, on the authority of the celebrated case of *Sturges v Bridgeman*,[7] that what constituted a common and ordinary use of land was to be judged having regard to the character of the locality.[8] A further rule illustrated by *Sturges* was that coming to a nuisance was no defence.[9] Neither was it a defence that the defendant's activity did not amount to a nuisance until the claimants' land was built on or its use had changed. The rule that coming to a nuisance was not a defence was confirmed recently by the Supreme Court in *Lawrence v Fen Tigers Ltd*.[10]

Lord Leggatt then applied the law, which he had summarised, to the facts of the case. He was of the view that it was beyond doubt that the viewing and photography which took place from the Tate building caused a substantial interference with the ordinary use and enjoyment of the claimants' properties.[11] Furthermore, such use could not be said

3 Ibid [21].
4 Ibid [23].
5 Ibid [25].
6 Ibid [27].
7 (1879) 11 Ch D 852 (CA).
8 [2023] UKSC 4, [38].
9 Ibid [42].
10 [2014] UKSC 13.
11 [2023] UKSC 4, [48].

to be a necessary, or ordinary, incident of operating an art museum.[12] Hence, the Tate could not rely on the principle of give and take and argue that it sought no more toleration from its neighbours for its activities than they would expect the Tate to show for them.

Lord Leggatt then addressed the decisions of the trial judge and the Court of Appeal. Both courts had rejected the claimants' claim for entirely different reasons. Lord Leggatt stated that the lower courts had erred under three heads.[13]

THE DECISION AT FIRST INSTANCE

As far as the decision of the trial judge was concerned, Lord Leggatt stated that the former had erred by framing the question which he had to decide, as to whether the Tate was making an unreasonable use of its land by operating the viewing gallery as it did.[14] Instead the trial judge should have ascertained whether it was a common and ordinary use. Lord Leggatt stated that having asked himself the wrong question, the answer, unsurprisingly, was that the operation of a viewing gallery was not an inherently unreasonable activity in the neighbourhood.[15] Nowhere did the judge consider whether the operation of a viewing gallery was necessary for the common and ordinary use and occupation of the Tate's land. Lord Leggatt stated that, had the trial judge done so, he would have been bound to conclude that, as in *Bamford v Turnley*,[16] the Tate was not using its land in a common and ordinary way, but in an exceptional manner.

Lord Leggatt then stated that the trial judge had applied the law incorrectly, in considering the impact of the Tate's activities on the ordinary use and enjoyment of the claimants' flats. Lord Leggatt addressed separately the judge's reasoning in relation to (a) the sensitivity of the flats and (b) the availability of protective measures.[17]

As far as (a) was concerned, Lord Leggatt agreed with the judge that the glassed design of the claimants' flats and their sensitivity to inward view was a relevant factor. It was relevant to the visual intrusion which the occupants could be expected to tolerate.[18] However, the judge went wrong in how he analysed that question. Critically, the judge did not distinguish between two different types of argument, one of which was valid, and the other which was not.

---

12 Ibid [50].
13 Ibid [53].
14 Ibid [54].
15 Ibid [55].
16 (1862) 3 B & S 66, 122 ER 27.
17 [2023] UKSC 4, [56].
18 Ibid [61].

As far as the valid argument was concerned, the trial judge was plainly right to say that floor to ceiling windows were an advantage which came at a price, in terms of privacy.[19] The fact that the property had been designed in such a way that made the occupants particularly vulnerable to inward view could not increase the liability of neighbours. Lord Leggatt gave the hypothetical example of another block of buildings, of similar height, being erected on the site of the Blavatnik Building (where the viewing gallery was currently situated) in such circumstances that the occupants of these flats could see straight into the claimants' living accommodation, causing annoyance to the claimants.[20] In these circumstances, if the occupants of the new flats were doing no more than making a normal use of their own homes, and showing as much consideration for their neighbours as they could reasonably expect their neighbours to show for them, the claimants could not have complained of nuisance. Such a situation would be analogous to the facts of *Southwark LBC v Mills*,[21] where the claimants had to put up with the noise, which was incidental to the ordinary use and occupation of neighbouring flats, despite the considerable annoyance, resulting from the fact that flats had been constructed without adequate sound insulation. Lord Leggatt added that, in the same way, in accordance with the principle of reciprocity, each flat owner (in the example given) would have to put up with being visible to their neighbour. That would be required by the rule of 'give and take, live and let live'.

Lord Leggatt went on to state that it did not follow that where a person was using land, 'not in a common and ordinary way, but in an exceptional manner', it was a defence to argue that a neighbour would not have a material inconvenience, were it not for the fact that the neighbour occupied an 'abnormally sensitive' property.[22] He stated that the nature and the extent of the viewing of the claimants' flats went beyond anything which could reasonably be regarded as a necessary or natural consequence of the common and ordinary use and occupation of the Tate's land.[23] That could not be regarded as a common or ordinary use of land.[24]

Lord Leggatt then addressed issue (b): that is, whether the claimants could have adopted relevant measures to protect themselves from being overlooked from people on the viewing gallery. The trial judge had stated that, as far as the visual intrusion of the claimants' homes

19 Ibid [62].
20 Ibid [63].
21 [2001] 1 AC 1 (HL).
22 [2023] UKSC 4, [65].
23 Ibid [74].
24 Ibid [75].

was concerned, if the interior of a person's home could be seen from the windows of houses across the street, and the occupants wished to avoid being seen, it was for them to draw their blinds, or take other remedial measures.[25] However, Lord Leggatt stated that in circumstances where the claimants were doing nothing other than occupying and using their flats in a common and ordinary way, and in accordance with the ordinary habits of a reasonable person, it was no answer for someone who interfered with that use by making an exceptional use of their own land to say that the claimants could protect themselves in their own homes by taking remedial measures.

THE DECISION OF THE COURT OF APPEAL

Lord Leggatt then addressed the decision of the Court of Appeal. He stated that the sole reason why the Court of Appeal did not find the Tate liable in nuisance was that liability in nuisance did not extend to overlooking.[26] Lord Leggatt agreed with that proposition but disagreed with the Court of Appeal's view that the claimants' claim concerned 'overlooking'. Lord Leggatt then emphasised that the claimants' complaint was not the fact that their flats were overlooked from the Blavatnik Building.[27] Rather, they complained about the use which had been made of the top floor by the Tate. The Tate had actively invited members of the public to visit and look out from the viewing gallery in every direction, including the claimants' flats situated about 30 metres away, without interruption, for the best part of the day. That constituted a nuisance. Lord Leggatt added that the notion that visual intrusion could not constitute a nuisance was not supported by precedent.[28]

Lord Leggatt then addressed three policy reasons which the Court of Appeal advanced for rejecting the claimants' appeal.

The first was that the Court of Appeal was of the opinion that it would be difficult to apply an objective test for deciding if there had been a material interference with the amenity value of the land.[29] In rejecting that ground, Lord Leggatt stated that intrusive viewing was no more subjective, or harder to judge, than any other forms of nuisance.[30] There was nothing peculiar about assessing whether visual intrusion amounted to a nuisance.

---

25 Ibid [84].
26 Ibid [89].
27 Ibid [92].
28 Ibid [104].
29 Ibid [106].
30 Ibid [108].

The second matter of policy raised by the Court of Appeal was that planning laws would be a better medium for controlling 'inappropriate looking' than the common law of nuisance. However, Lord Leggatt endorsed Lord Neuberger's *dictum* in *Lawrence v Fen Tigers* to the effect that there was no basis, in principle, for the proposition that the planning regime 'cut down' private rights.[31]

The third policy reason advanced by the Court of Appeal concerned the issue of privacy. The Court of Appeal stated that there were other laws which bore on privacy. An extension of the law could only be made by Parliament rather than by the courts. However, Lord Leggatt stated that that view assumed that applying the law of nuisance to the facts of the instant case would require an extension of the law.[32] That was a wrong assumption. No new privacy laws were required to deal with that complaint.

Lord Leggatt then addressed the relevance of public interest in the decisions of the lower courts. Both lower courts were influenced by what they perceived to be the public interest in the use made of the viewing gallery.[33] However, public interest was not a factor that fell to be addressed when the court was deciding whether the use which was being made of the viewing gallery amounted to a nuisance. It was a factor one should apply when deciding whether to grant an injunction or an award of damages after it had been decided that a nuisance existed.[34] Lord Leggatt added that the point of private nuisance was to protect equality of rights between neighbouring occupiers to the use and enjoyment of their own land when those rights conflicted. In deciding whether one party's use had infringed the other's rights, the public utility of the conflicting uses was not relevant.[35]

Lord Leggatt concluded that the use of the Tate's viewing gallery constituted a nuisance in law.[36]

THE DISSENTING JUDGMENT

Lord Sales (with whom Lord Kitchen agreed) gave a dissenting judgment. He stated that the instant case raised two questions.

The first was, in principle, whether it was possible to conclude that a private nuisance existed, in the case of residential property, by reason

31 Ibid [110].
32 Ibid [111]–[112].
33 Ibid [114].
34 Ibid [120].
35 Ibid [121].
36 Ibid [133].

of visual intrusion consisting of people looking into the living areas of the claimants' property?[37]

The second question was that, if that was possible, had the claimants established that there was an actionable private nuisance by reason of the visual intrusion which they had experienced?

Lord Sales stated that visual intrusion into someone's domestic property was capable of amounting to a nuisance.[38] There was no good reason to rule out the claimants' claim as a matter of principle.[39]

In relation to the second question, as to whether a nuisance existed in the instant case, Lord Sales stated that the application of the 'give and take' principle, as a way of modulating and reconciling property rights of neighbouring landowners, was particularly important where the issue was visual intrusion, or overlooking.[40] He stated that he saw no good reason why one should leave out of account reasonable self-help measures (such as the provision of blinds and curtains) which might be available to the person complaining about visual intrusion.[41] In turn, it was possible for the Tate to reduce the impact from the viewing platform on the claimants' property by closing it at certain times, putting up notices, and taking similar steps.[42]

Lord Sales acknowledged that 'coming to a nuisance' was no defence, and that the 'give and take' principle was an objective one, which was to be applied in the light of the nature of the neighbourhood in which the relevant properties were located. He stated that there were sound reasons why the law adopted an objective approach in the context of the relevant locale.

The first of these reasons was that elevating the question of whether the defendant had acted in accordance with the existing common and ordinary use of land in the locality into the ultimate test for nuisance would seriously distort the law of nuisance.[43] Such an exclusive focus placed excessive weight on one side of what was an inextricably two-sided relationship. This would mean that, if a defendant's use of land was outside such use, the claimant would simply require to prove that the defendant's use of land had an unwelcome impact on the claimant's use of their own land.

The second reason was that a claimant landowner and a defendant landowner might each wish to use their property in ways which were not in themselves common, according to the standards of the locale,

37 Ibid [134].
38 Ibid [179].
39 Ibid [204].
40 Ibid [212].
41 Ibid [214].
42 Ibid [216].
43 Ibid [227].

and the test to govern any conflict between these two uses had to be capable of accommodating such situations, in a just manner.[44]

Thirdly, to make the exposure of the defendant depend on common and ordinary usage of its land was too conservative as regards the development of land and conflicted with the general policy of the law that a landowner should be free to use their land as they wished.[45]

Fourthly, whereas questions of common and ordinary usage of land by a defendant might be central in working out the application of an objective standard of reasonableness in a locale, they were not, in themselves, capable of providing a solution across the whole range of cases with which the law of nuisance was required to deal.[46] It was necessary to have regard to a more general principle of objective reasonableness. Lord Sales added that the Tate's use of its land by operation of the viewing gallery was not a common and ordinary use of the land in the locale.[47] However, that factor was not sufficient to render the Tate liable to a claim in nuisance by any neighbouring landowner who could say that the resulting interference with their interests was significant or substantial. The claimants' use of their land, by adopting an unusually open form of design for residential living in the relevant urban locale and using the winter gardens as they did, was not a common and ordinary use of land in that locale.[48] Therefore, the claimants were not in a position, for their part, simply to claim that the Tate was obliged to moderate the use of its land, according to the objective standards of reasonableness, applicable in that locale.

Lord Sales went on to state that, fifthly, an objective test of reasonable reciprocity and compromise was clear and workable.[49]

The sixth point which Lord Sales made was that a test which was based on the common and ordinary use by the defendant, was contrary to the way the test was formulated in the modern authorities.[50] The rule of 'give and take, live and let live' was a general test of objective reasonableness and had been approved in recent cases of the highest level.

Lord Sales then addressed the decision of Mann J. The latter had found that the law of nuisance could apply in cases of invasion of privacy by visual intrusion in relation to residential property.[51] He also had

---

44 Ibid [229].
45 Ibid [231].
46 Ibid [232].
47 Ibid [237].
48 Ibid [238].
49 Ibid [240].
50 Ibid [243].
51 Ibid [256].

found that the Tate was making reasonable use of its land.[52] Lord Sales stated that Mann J had assessed the standards of privacy which were to be expected in the locale at Neo Bankside and had concluded that owner/developers of dwellings designed with heightened vulnerability to external gaze in that locale could not complain. Lord Sales went on to state that Mann J had found that the atypical design of the flats, in the context of the standards of privacy which were reasonably to be expected in that locale, was a relevant factor in determining the overall reasonableness, as between parties, according to an objective assessment.[53] The latter had concluded that it would be wrong for the self-induced incentive to gaze into the flats, associated with their exceptionally open design, to create liability in nuisance.[54] Mann J had also concluded that, as far as the protection of the claimants' privacy was concerned, it was reasonable to expect the claimants to 'protect their own interests' to some degree.[55]

Lord Sales then addressed the decision of the Court of Appeal. That court had criticised Mann J's judgment on two grounds.[56] First, the latter had been wrong to conclude that the claimants were required to take self-help measures to prevent the visual intrusion of their flats. Secondly, the court had held that the claimants were using their flats in a perfectly normal fashion, as homes. The trial judge's approach in balancing these interests against those of the Tate was contrary to the principles of private nuisance.

However, Lord Sales stated that the Court of Appeal's criticisms of Mann J's judgment were wrong.[57] Lord Sales stated that the Court of Appeal had given insufficient weight to the reasonable interest of the Tate in making use of its own property as it wished by operating a viewing gallery, which Mann J had found to be reasonable, when assessed by reference to the locality.[58] The Court of Appeal had found that the viewing gallery was not necessary for the common and ordinary use of the Tate's property. However, Lord Sales stated that there was no good reason why the give-and-take test should be weighed against one of the competing property owners in such a way. The Court of Appeal had distorted the give-and-take principle by setting the interests of the claimants to use their property as was *reasonable* against a test which would require the Tate's use of its property to satisfy the higher standard of being *necessary*. Lord Sales held that Mann J's approach

52 Ibid [257].
53 Ibid [261].
54 Ibid [262].
55 Ibid [263].
56 Ibid [265].
57 Ibid [269].
58 Ibid [270].

to the give-and-take principle was correct.[59] Property owners in that part of London expected to be overlooked, and it was normal to expect people to use curtains and the like to limit the annoyance that might be caused. Mann J had found that the viewing gallery would not have caused a nuisance if the claimants' property had been used in such a way that did not involve heightened sensitivity to visual intrusion. Lord Sales stated that the owners of the land at Neo Bankside chose to develop it by building striking buildings of architectural distinction, which was likely to attract attention and the gaze of strangers.[60] In assessing what was the reasonable balance to strike between the competing interests and property rights of the claimants and the Tate, in the context of the particular neighbourhood and in the light of the particular nuisance alleged (ie visual intrusion), the trial judge had been entitled, in the circumstances, to have regard to the availability of self-help measures, which it was not unreasonable to expect the claimants to take.[61] Lord Sales added that the Tate could not turn the operation of the viewing gallery into a nuisance, by reason of the development of their own property, according to a design which was out of line with the norm for that area.[62]

In conclusion Lord Sales stated that he would have dismissed the appeal.[63]

COMMENT

The Supreme Court was required to address two main substantive issues.

The first was whether the act of being overlooked by individuals standing on the defendant's viewing gallery, which adjoined the claimants' flats, could rank as a nuisance in law.

The second was whether the fact that the claimants occupied flats, the external walls of which were constructed entirely of glass thereby allowing visitors standing on the viewing gallery to stare into the flats, rendered the claimants' flats 'oversensitive', thus depriving the claimants a remedy by way of the law of nuisance.

As far as the first issue was concerned, there was, indeed, scanty authority to the effect that visual intrusion could form the basis of a nuisance action. The vast majority of nuisance cases have involved unreasonable interference with the claimant's land by smoke, fumes, odours, flooding, noise, vibrations and so on. However, as Newark

59 Ibid [271].
60 Ibid [272].
61 Ibid [273].
62 Ibid [278].
63 Ibid [280].

observed in his seminal article on nuisance,[64] the boundaries of the tort of nuisance are blurred. Indeed, neither judge nor academic has been able to offer a comprehensive and clear definition of what constitutes a nuisance. Furthermore, in the Court of Appeal case of *Thompson-Schwab v Costaki*,[65] where the court held that the sight of prostitutes and their clients entering and leaving premises in the vicinity of the claimant's house could constitute a nuisance, Lord Evershed MR stated that 'the forms which activities constituting actionable nuisance may take are exceedingly varied'. He added that they were not capable of precise or close definition. In short, the list of the various ways in which the claimant's enjoyment of their land can be adversely affected is not closed. Indeed, Lord Leggatt stated that anything short of direct physical invasion of the claimant's land could constitute a nuisance. Therefore, there was no doctrinal reason to preclude the court from deciding that unreasonable visual intrusion of the claimants' flats could rank as a nuisance. Indeed, in *Watt v Jamieson*[66] Lord Cooper stated that *any* type of use of the defendant's property which subjected adjoining proprietors to substantial annoyance was *prima facie* not a reasonable use and, therefore, capable of being a nuisance. Therefore, Lord Leggatt's deciding that visual intrusion could rank as a nuisance did not fall foul of any principle either in English or Scots law. However, the decision does take the law further forward.

As far as the second issue is concerned, it is well established that the defendant is liable in nuisance only if the use of their land is unreasonable.[67] For Lord Leggatt the principles of reciprocity and equal justice underpinned the concept of unreasonableness. These principles were articulated, in terms of the law of nuisance, by the rule that an occupier of land could not complain if the use of the land which was being interfered with was not an *ordinary* use. Conversely, even if the defendant's conduct substantially interfered with the ordinary use of the claimant's land, no action in nuisance would lie if the defendant's activity was no more than the ordinary use of the land. In the instant case, Lord Leggatt held that, whereas the Tate could have been using the viewing gallery reasonably, the existence of the gallery was not a common and ordinary use of land and, therefore, *prima facie* capable of constituting a nuisance. However, having established that the viewing gallery was not a common and ordinary use of land, one was then required to determine the nature of the claimants' use of land and juxtapose that use with that of the defendant. For Lord Leggatt the fact that the external walls of the claimants' flats were constructed of glass

64 F H Newark, 'The boundaries of nuisance' (1949) 65 Law Quarterly Review 480.
65 [1956] 1 WLR 335 (CA), 338.
66 1954 SC 56, 58.
67 *Baxter v Camden LBC (No 2)* [2001] QB 1 (CA).

did not take that use out of that which was common and ordinary. In short, such use was not oversensitive, thereby depriving the claimants a remedy by way of the law of nuisance.

Traditionally, however, the courts have addressed the question of whether the claimant's use of land is oversensitive and thereby unable to be protected by an action in nuisance, without comparing the defendant's use or user of land with that of the claimant.[68] Lord Leggatt's approach in *Fearn* takes the law further forward, by requiring a comparison to be made of use made of the claimants' property and the use made of the land of the defendant, in terms of that which ranks as common and ordinary. According to Lord Leggatt, if the defendant's use of land deviates to a greater extent from the norm than that of the claimant, in that respect, the latter can succeed in a nuisance action. In *Fearn* the defendant's use of land deviated from the norm (ie that which ranked as common and ordinary) to a greater extent than did that of the claimants. Therefore, it automatically followed that the defendants use of the gallery constituted a nuisance. However, the author would argue that it may often be difficult, first, to determine whether any use of the defendant's land deviates from the norm, in terms of a given locality, and, secondly, to weigh, as it were, such a use against that of the claimant, in terms of the law of nuisance. The author would, therefore, readily endorse the dissenting view of Lord Sales, to the effect that elevating the question of whether the defendant had acted in accordance with the existing common and ordinary use of land in the locality, as to be the ultimate test for nuisance, would seriously distort the law of nuisance, in that, if the defendant's use of land lay outside the norm of that which was common and ordinary, a claimant would only be required to show that the defendant's use had an unwelcome impact on the claimants' land. Furthermore, Lord Sales' endorsement of the general test of reasonableness, as the test for liability in nuisance, conforms to the traditional approach which has been adopted by the courts, is more conducive to clarity and also allows the law to develop both coherently and incrementally, by considering the changing requirements and expectations of society.

Finally, the social utility of the defendant's activities has traditionally been considered as going to the question of whether a nuisance exists.[69] However, in *Fearn* Lord Leggatt was of the view that the fact that the

68 *Robinson v Kilvert* (1888) 41 Ch D 88 (CA); *Heath v Brighton Corporation* (1908) 24 TLR 414; *Bridlington Relay Ltd v Yorkshire Electricity Board* [1965] Ch 436 Ch D); *Network Rail Infrastructure Ltd (formerly Railtrack) v Morris* [2004] EWCA Civ 172.

69 *Harrison v Southwark and Vauxhall Water Co* [1891] 2 Ch 409 (Ch D); *AG v Hastings Corp* (1950) 94 Sol J 225 (CA); *Lawrence v Fen Tigers Ltd* [2014] UKSC 13, [185] (Lord Carnwath).

Tate had public utility was an irrelevant factor in the court determining whether the viewing gallery was a nuisance. Rather, public utility had relevance solely in relation to the remedy, if any, which fell to be awarded against the defendant. In this respect, Lord Leggatt followed the decision of Buckley J in *Dennis v MoD*.[70] However, Lord Pentland refrained from expressing a view as to whether *Dennis* represented the law of Scotland in *King v Lord Advocate*.[71] The author would suggest, however, that *Dennis* does not, and that public interest should be considered by the court when it decides whether a nuisance exists.

70 [2003] EWHC 793 (QB), [2003] Env LR 34.
71 [2009] CSOH 169, [17].

Articles

Commentaries and Notes