ISSN: 2753-5746

CAMBRIDGE LAW REVIEW

VOLUME VII
ISSUE 2
AUTUMN 2022

Editor-In-Chief

Andreas Samartzis

EDITOR-IN-CHIEF'S INTRODUCTION TO THE AUTUMN ISSUE OF VOLUME VII OF THE CAMBRIDGE LAW REVIEW

It is with great pleasure that I present the Autumn Issue of Volume VII of the Cambridge Law Review. Another academic year has come to an end, over which our journal has received a significant number of interesting submissions.

In this issue's first article, Scott Morrison analyses the legal, economic, and political implications of Brexit for Northern Ireland in light of the Northern Ireland Protocol. The author considers that the Protocol has failed to strike the right balance between nationalists and unionists. In favouring the nationalist position of avoiding a hard border between the Republic of Ireland and Northern Ireland, the Protocol has disrupted the movement of goods between Northern Ireland and Britain. It has done so by imposing customs duties on all goods subject to commercial processing, as well as to goods which are at risk of being sold in the EU single market. As the local economy relies primarily on trade with Britain, this development has had an alarming impact on the economy of Northern Ireland. The author examines two solutions to the particular problem of the disruption of the food supply, one involving the UK aligning with EU regulations on food products, the other advocating for a precise definition of the class of products at risk of being sold in the EU single market. The author considers the former to be the most effective with respect to resolving the disruption of the food supply, but concedes that it is politically unfeasible under the present circumstances. The article then turns to the legal enforcement of the Northern Ireland Protocol. The author here expresses some doubt on whether the preliminary reference procedure to the Court of Justice of the European Union and the possibility of opening infringement proceedings for violation of the Protocol will result in the effective resolution of legal disputes, given the UK's dualist legal system and the UK government's statements about the prospect of disregarding international law. Finally, the article examines the political unrest following the adoption of the Protocol and the dissatisfaction—which has led to litigation—around the fact that the Northern Irish Assembly is disabled from having a voice on the continuing validity of the Protocol until 2024.

In "'Legitimate' Protest in European Human Rights Law: A Critical Reconstruction", Mythili Mishra studies the construction of 'legitimate' protest in

European human rights law. Her article uses the jurisprudence of the European Court of Human Rights to understand and evaluate what kinds of protest the Court legitimises, and what kinds it does not. It does so on the basis of three ideas, namely responsibility, disruption, and offence. The author argues that these three fundamental strands come together to construct the Court's account of 'legitimate' protest. This account is also reconstructed in the article through a critical evaluation of the Court's justifications, which enable one to interrogate the Court's judgments and criticise them for inadequately protecting the right to protest. The article concludes with observations about what its findings mean for the protection of human rights and democracy. In particular, it posits that the Court offers only limited or no protection to protestors who do not fit a certain model. This practice constitutes, in the author's view, a threat to democracy.

In "A Critical Analysis of the Scottish Government's Draft Gender Recognition Reform (Scotland) Bill and its Adherence to the UN Convention on the Elimination of All Forms of Discrimination against Women", Esther Hodges examines the Scottish government's draft Gender Recognition Reform (Scotland) Bill. The draft Bill aims to streamline the process for those seeking to obtain a Gender Recognition Certificate and so amend the sex on their birth certificate to the gender with which they identify. Its proposed reforms have attracted significant opposition. Drawing on qualitative analysis of submissions to the draft Bill's second public consultation, this article argues that opposition is typically based on a reductive, classical sociological conceptualisation of gender, which understands gender as an immutable binary ordained by nature and contends that trans women are not women. By making it easier for trans women to gain legal recognition for the gender with which they identify, those opposing the draft Bill on these grounds therefore argue that its reforms put the rights and freedoms of cis women at risk. The article explores this contention by critically analysing the draft Bill's adherence to the Convention on the Elimination of All Forms of Discrimination against Women (CEDAW). Setting analysis against a framework of two of the CEDAW's most relevant articles and its General Recommendation 28, it argues that the draft Bill is demonstrably in adherence with CEDAW because of its efforts to reduce discrimination against trans women through means which in no way increase the risk of discrimination against cis women. Finally, the author argues that the draft Bill, and indeed CEDAW, could go further in their efforts to reduce discrimination faced by trans women by reducing their evidential reliance on binary conceptualisations of gender. In so doing, the Bill could encourage greater feminist and queer coalitional work, discouraging efforts to pit women's rights against those of trans people.

The issue's fourth article, titled "An Assessment of the Effectiveness of the Unfair Prejudice Remedy in UK Company Law: How can we Guarantee Appropriate Judicial Discretion?" brings attention to section 994 of the Companies

Act 2006, which enables members of a company to petition the court for a remedy in respect of conduct by other members that unfairly prejudices their interests. The author, Ziyuan Li, argues that, although the court's open-ended interpretation of s 994 provides a reliable safeguard for the minority shareholders' interests, it may indirectly encourage opportunistic behaviour leading to abusive unfair prejudice actions. The rapidly growing number of s 994 petitions have led to this type of proceeding becoming more burdensome, thereby increasing the financial and time burden on both the petitioner and the court. Moreover, the expansive discretion has resulted in an overlap in jurisdiction between s 994 petitions, which traditionally represent personal relief, and derivative claims, which represent corporate relief. The author's view is that this development opens the floodgates for minority shareholders to bring malicious claims to interfere with the affairs of the company. As a consequence, the unfair prejudice remedy regime may run counter to the objectives of 'efficiency' and 'fairness' in the area of shareholder remedies law. In light of this background, the article explores how the effectiveness of the s 994 petitions may be improved. It does so by proposing a guiding framework for the construction of appropriate judicial discretion that better balances shareholder protection and corporate autonomy.

In "A Necessary Shift from Shareholder Primacy toward Stakeholder-Conscious Governance in Light of Corporate Social and Environmental Responsibility", Sophie Treacy engages with the effect of climate change on corporate governance. The author identifies a tension between the traditional model of shareholder primacy, with its priority to profit-maximising activities, and the need for a sustainable and socially responsible governance. To address this tension, the article argues that, at the level of doctrine, corporate purpose is undergoing a paradigm shift from strictly shareholderist to stakeholder-conscious governance, prompted by a growing number of social and environmental exigencies. The author supports this statement by examining the origins and normative legitimacy of shareholder primacy, along with the extent to which shareholderist governance can be reconciled with activities of corporate social responsibility. The article concludes that shareholder primacy is teetering on the brink of collapse, as the climate crisis demands corporate purpose to evolve toward a much more holistic, stakeholder-conscious model of governance.

In his article, "The Strange Saga of Compensatory Taxes: Charting a Way Out of India's Maze of Doctrinal Uncertainty", Rishabh Jain reviews the 'doctrine of compensatory taxation' ('DCT') – that is, the levy of supposedly non-restrictive taxes by States on account of facilitation of trade, commerce, and intercourse ('TCI') – in Indian constitutional law. The article analyses the rise and fall of the DCT to reveal underlying conflicts between its two referents which explain the 'maze of doctrinal uncertainty' around it. It then argues that the DCT was rightly rejected, but that the conceptual confusions introduced by it have not been fully

extirpated and have even figured in some of the grounds used to reject the DCT. In the process, the article engages with some of the rough edges of Indian constitutional jurisprudence, such the conflation of two different and contrary doctrines under the common label of DCT, the rejection of the older 'direct and immediate impact' doctrine due to conflation with DCT, and the mislocation and misapplication of the fee-tax distinction. The author concludes by regarding the strange saga of the DCT as a warning against over-emphasis in TCI jurisprudence on textual factors to the exclusion of conceptual (particularly economic and logistical) ones.

The issue concludes with three case comments. In "Justice Shortchanged? Redrawing the Ethical Boundaries of Lifted Judgments Following *Crinion v IG Markets Ltd* [2013] EWCA Civ 587", Shen-Way Chong makes the case for the imposition of a duty on the bench not to plagiarise the language of the victorious litigant—what he refers to as the practice of issuing lifted judgments—and to advocate for a "functional approach" to deal with instances of unbridled judicial copying. By referring primarily to the leading Court of Appeal decision in *Crinion v IG Markets*, the author seeks to illustrate the paradox between lifted judgments and the ethics espoused in the Guide to Judicial Conduct.

"Assumptions of Irresponsibility: Liability for Omissions following *Tindall v Chief Constable of Thames Valley*" analyses the present state of negligence liability in English tort law as set out in the recent case of *Tindall v Chief Constable of Thames Valley*. Despite recent landmark decisions regarding acts and omissions, the boundaries of the distinction between the two remain to be fully explored. Following the decision in *Tindall*, the author, Sam Pearce, suggests that a temporary conferral of a benefit must always fall to be classified as an omission. The article then argues that, for a claimant to establish that a defendant has assumed a responsibility to them, it must first be shown that the defendant has a relationship with the claimant that is sufficiently distinguishable from the general public. It is the lack of such a relationship that prevented the claimant in *Tindall* from successfully arguing that the police had assumed a responsibility to all road users. This commentary concludes that Tindall further elucidates key duty of care principles under the law of negligence, whilst also highlighting important questions that will require clarification from the courts in the future.

The issue's final case comment, "Dichotomy between Jurisdiction and Admissibility: Illuminating the Twilight Zone – *BTN v BTP* [2021] 1 SLR 276", authored by Joel Soon, reviews the Singapore Court of Appeal's judgment in *BTN v BTP*. According to the author, the significance of this judgment lies in that it affirmed that the tribunal versus claim test, which was introduced in the Singapore Court of Appeal's earlier judgment in *BBA v BAZ*, continues to apply to determine whether issues go towards jurisdiction or admissibility. Notwithstanding the strong impetus for drawing a dichotomy between jurisdiction and admissibility, the

dichotomy's usefulness is called into question where issues defy easy classification. The inflexibility perpetuated by the dichotomy has led, according to the author, to the emergence of a twilight zone. The central thesis of the case comment is that the dichotomy may be of limited usefulness in certain areas in the law of arbitration, although, admittedly, the Singapore courts are stuck between a rock and a hard place since alternatives to the dichotomy come with shortcomings of their own.

I wish to thank our team of Senior, Associate, and International Editors for their work and dedication during this period.

Andreas Samartzis
Editor-in-Chief

Cambridge Law Review Editorial Board

Honorary Board
Judge Awn Al-Khasawneh
Michael Blair KC
The Honourable Sir Jeremy Cooke KC
Jern-Fei Ng KC
Judge Hisashi Owada
Justice Anselmo Reyes SC
Professor Malcolm Shaw KC

Editor-in-Chief
Andreas Samartzis

Managing Editor
Jared Foong

Senior Editors
Durga Agarwal
Barry Dynkin
Fred Halbhuber
Vandita Khanna
Folakunmi Pinheiro
Quentin Benedikt Schäfer
Vinamrata Shrivastava
Isabelle St-Hilaire
Christopher Symes
Bagavathy Vennimalai

Associate Editors
Frederick Abu-Bonsrah
Daniel Pereira Campos
Marcus Choy
Cameron Coleman
Rory Craig
Wednesday Eden
Aleksandra Fiutowska
Andrew Furlan
Ana Biderman Furriela
Luke Gallagher
Kartik Suramanian Kann
Dushyant Kishan Kaul
Katja Kreft
Nishanth Vasanth Kumar
Neelasha Nemani
Leo Pang
Jin Qin
Harry Sanderson
Jai Sanyal
Apoorva Satapathy
Tooba Shahzad
Yigit Anil Tarman

International Editors
Oluwasoji Temidire Ayorinde
Obefemi Awolowo University
Yagmur Hortoglu
New York University
Ismini Mathioudaki
Panteion University
Isha Prakash
Government Law College
Orlaith Rice
University College Dublin
Snehil Kunwar Singh
National Law School of India University
Aakriti Tripathi
Jindal Global Law School
Marcel Zernikow
University Paris I Pantheon-Sorbonne

Table of Contents

The Northern Ireland Protocol: A Long-Term Solution to the Economic, Legal, and Political Impacts of Brexit on Northern Ireland? 1
Scott Morrison

'Legitimate' Protest in European Human Rights Law: A Critical Reconstruction 28
Mythili Mishra

A Critical Analysis of the Scottish Government's Draft Gender Recognition Reform (Scotland) Bill and its Adherence to the UN Convention on the Elimination of All Forms of Discrimination against Women 56
Esther Hodges

An Assessment of the Effectiveness of the Unfair Prejudice Remedy in UK Company Law: How can we Guarantee Appropriate Judicial Discretion? 72
Ziyuan Li

A Necessary Shift from Shareholder Primacy toward Stakeholder-Conscious Governance in Light of Corporate Social and Environmental Responsibility 103
Sophie Treacy

The Strange Saga of Compensatory Taxes: Charting a Way Out of India's Maze of Doctrinal Uncertainty 119
Rishabh Jain

Justice Shortchanged? Redrawing the Ethical Boundaries of Lifted Judgments Following Crinion v IG Markets Ltd [2013] EWCA Civ 587 134
Shen-Way Chong

Assumptions of Irresponsibility: Liability for Omissions following Tindall v Chief Constable of Thames Valley 147
Sam Pearce

Dichotomy between Jurisdiction and Admissibility: Illuminating the Twilight Zone BTN v BTP [2021] 1 SLR 276 153
Joel Soon

Cambridge Law Review (2022) Vol VII, Issue 2, 1–27

The Northern Ireland Protocol: A Long-Term Solution to the Economic, Legal, and Political Impacts of Brexit on Northern Ireland?

SCOTT MORRISON*

ABSTRACT

The Northern Ireland Protocol is the solution agreed by the European Union and the United Kingdom to the unique problems arising as a result of Brexit on the island of Ireland. The Protocol preserves the "soft" border between Ireland and Northern Ireland and ensures the all-Ireland economy will remain undistorted. In contrast, trade between Northern Ireland and Great Britain is now subject to onerous customs duties and tariffs, as a result of Northern Ireland's *de facto* continuing membership of the European single market. This article examines the key aspects of the Protocol—economic, legal, and political—and seeks to demonstrate that the Protocol is far from a perfect solution to the situation in Northern Ireland.

Keywords: Northern Ireland, European Union, United Kingdom, constitutional law, Brexit

I. INTRODUCTION

A. OVERVIEW

"'History,' Stephen said, 'is a nightmare from which I am trying to awake.'"[1]

* LLM Candidate (Durham University), LLB (Durham University). I am grateful to the anonymous reviewers for their comments on earlier drafts. Any errors that remain are my own.

[1] James Joyce, *Ulysses* (Oxford World's Classics 2008) 34.

As of 2022, it would appear that the nightmare of history has once again descended on Northern Ireland. Ironically, the Northern Ireland Protocol,[2] one of the causes of the new troubles the region finds itself in, was designed to prevent further violence.[3] The Protocol is a unique solution to a unique problem arising from Brexit, keeping Northern Ireland *de facto* aligned with the EU's single market and customs union in the interests of avoiding a hard border on the island of Ireland and preserving the Northern Irish peace process.

This introduction consists of two parts. The first will explain the Brexit process and how the Protocol became a necessary solution. The second will explain the context of the problems in Northern Ireland and why the Protocol was needed to avoid reigniting tensions between its communities.

The United Kingdom of Great Britain and Northern Ireland (UK) voted to leave the European Union (EU) in a referendum on 23 June 2016.[4] On 29 March 2017, the then UK Prime Minister Theresa May formally notified the then President of the European Council, Donald Tusk, that the UK would invoke Article 50 of the Treaty on European Union (TEU), so beginning a two-year process of negotiations which would culminate in the UK leaving the EU on 29 March 2019.[5] Two years of negotiations aimed at concluding a withdrawal agreement followed.[6] The leaders on the European side agreed to a withdrawal agreement on 25 November 2018, whereas the UK Parliament did not vote for the agreement, leading to two extensions of the Article 50 deadline: the first date being 31 October 2019. Complicating this was Theresa May resigning in May 2019, an event brought about by May failing to get enough support in the UK Parliament for the deal she negotiated. Her successor as prime minister, Boris Johnson, negotiated a revised withdrawal agreement and called elections for 12 December 2019, which he won by a significant majority. The UK subsequently withdrew from the EU on 31 January 2020, which was the new date agreed upon after May's resignation.[7]

[2] 'Revised Protocol on Ireland and Northern Ireland included in the Withdrawal Agreement' (European Commission, 17 October 2019) <https://ec.europa.eu/info/publications/revised-protocol-ireland-and-northern-ireland-included-withdrawal-agreement_en> accessed 13 March 2021.

[3] Molly Blackall, 'Northern Ireland's first minister joins calls for calm after Belfast riots' *The Guardian* (London, 3 April 2021) <https://www.theguardian.com/uk-news/2021/apr/03/northern-ireland-secretary-calls-for-calm-after-belfast-riots> accessed 15 April 2021.

[4] Steven Erlanger, 'Britain Votes to Leave EU; Cameron Plans to Step Down' *The New York Times* (New York City, 23 June 2016) <https://www.nytimes.com/2016/06/25/world/europe/britain-brexit-european-union-referendum.html> accessed 14 February 2021.

[5] Stephen Castle, 'UK initiates 'Brexit' and Wades Into a Thorny Thicket' *The New York Times* (New York City, 29 March 2017) <https://www.nytimes.com/2017/03/29/world/europe/brexit-uk-eu-article-50.html> accessed 19 April 2021.

[6] Paul Craig and Gráinne de Búrca, *EU Law: Text, Cases and Materials UK Version* (7th edn, Oxford University Press 2020) 24.

[7] ibid 24–25.

The final Withdrawal Agreement[8] was agreed between the two parties, to which a specific protocol was added, the Northern Ireland Protocol.[9] The purpose of the Protocol is to avoid a hard border on the island of Ireland, protect the all-island economy and the Good Friday Agreement[10] in all its dimensions, and safeguard the integrity of the EU single market.[11] The protocol acknowledges the unique circumstances arising from the UK's withdrawal from the EU on the island of Ireland.

This article shall argue the Northern Ireland Protocol does not amount to a complete solution, particularly with respect to addressing challenges that arise in the UK's internal market and the potential incompatibility of certain aspects with the Good Friday Agreement. Therefore, this article shall adopt a critical stance towards the Protocol as needing reform to be a long-lasting solution to the problems caused by Brexit on the island of Ireland.

This article is divided broadly into three parts: the first part focuses on the Protocol's attempted solution to the economic impact of Brexit on Northern Ireland; the second part focuses on the Protocol's provisions for the continued role of EU law and the CJEU in Northern Ireland; and the third part focusses on the Protocol's attempted solution to the political problems brought about by Brexit and attempts to articulate and offer alternative solutions. Some contextual information relating to why the Protocol was agreed by the EU and the UK and an overview of the relevant legislation will be provided before the article focusses on the main economic, legal, and political issues.

B. CONTEXT OF THE IRISH BORDER QUESTION

A crucial issue in the Brexit negotiations was that of the border between Ireland and Northern Ireland, as Brexit meant Northern Ireland would no longer be part of the territory of the EU, while Ireland remained a member. This meant the need for border checks, customs, and so on, as the border between Ireland and Northern Ireland essentially became an EU external border; however, the problem was that because of Northern Ireland's history of conflict, it was felt that a hard border in this region would antagonise one of Northern Ireland's traditional communities, the Irish Catholics,[12] who see themselves as Irish, and wish for an all-Ireland independent state. The other traditional community are

8 Agreement on the withdrawal of the United Kingdom of Great Britain and Northern Ireland from the European Union and the European Atomic Energy Community [2019] OJ C-384I/01.

9 In international law, a protocol is a treaty that adds to or supplements a pre-existing treaty.

10 The Belfast Agreement' (*Northern Ireland Office*, 10 April 1998) <https://www.gov.uk/government/publications/the-belfast-agreement> accessed 13 March 2021.

11 The EU-UK Withdrawal Agreement' (*European Commission*, date unavailable) <https://ec.europa.eu/info/relations-united-kingdom/eu-uk-withdrawal-agreement_en> accessed 13 March 2021.

12 Also referred to as 'nationalists' or 'republicans'.

the Ulster Protestants,[13] who would prefer Northern Ireland to remain an integral part of the United Kingdom.

Boris Johnson may have dismissed the border dilemma as "the tail wagging the dog",[14] however, this matter was of profound significance to the citizens of Northern Ireland and Ireland, due to a long history of division, violence, and ethnic hatred between the island's two traditional communities. The most recent iteration of this centuries-long conflict only formally ended in 1998 with the signing of the Good Friday Agreement, though sporadic acts of violence continue to this day. One of the solutions to the conflict was to allow all persons born in Northern Ireland to choose Irish or British citizenship, or both, if they so wished. A Common Travel Area between the UK and Ireland had existed since the 1920s, meaning no customs or passport checks at the border, though the Troubles[15] meant crossing the border entailed checks from the British and Irish security forces. The withdrawal of most British troops from Northern Ireland in 2007 meant that crossing the border was seamless, a fact no doubt helped by both the UK and Ireland being EU member states. Brexit meant it was not possible to retain this *status quo*, and so all parties to the negotiations regarding the Withdrawal Agreement sought to ensure the UK's withdrawal from the EU would not result in a "hard border", that is to say, that it would not result in customs and passport checks.

Membership of the EU and its single market allows citizens, goods, services, and capital originating from the member states to move freely within EU territory.[16] Membership of the customs union means no tariffs or barriers to trade with other members. These two aspects of the Protocol ensure trade on the island of Ireland remains unfettered.

That Ireland and the United Kingdom were both members of the EU allowed what was formerly a hard and militarised border to become an invisible one.[17] During the Brexit negotiations, the UK government decided not to retain membership of the single market. As a consequence, it quickly became apparent that a radical new solution was needed in order to allow the UK to leave the single market and customs union; keep the Irish border free of physical infrastructure;

13 Also referred to as 'unionists' or 'loyalists'.

14 Ferghal Blaney, 'Boris Johnson slammed over 'tail wagging the dog' comments on Irish border Brexit issue' *Irish Mirror* (Dublin, 8 June 2018) <https://www.irishmirror.ie/news/irish-news/politics/boris-johnson-slammed-over-tail-12668455> accessed 14 February 2021.

15 The name commonly given to the most recent iteration of the conflict in Northern Ireland, which lasted from approximately 1968 until 1998.

16 Nikos Skourtaris, 'What's in an Irish Border? Brexit, the Backstop(s), and the Constitutional Integrity of the UK' (DCU Brexit Institute, February 24 2020) < https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3543514> accessed 14 February 2021.

17 ibid.

and maintain the integrity of the EU's single market and legal order.[18] The Protocol was designed by the EU and the UK to achieve all of those targets.

The next section of this introduction will provide an overview of the relevant legislation from the perspectives of the EU and the UK.

C. AN EXPLANATION OF THE IMPORTANT LEGISLATION

The main EU pieces of legislation, vis-à-vis Brexit and Northern Ireland, are the Withdrawal Agreement;[19] the Northern Ireland Protocol;[20] and the Trade and Cooperation Agreement (TCA).[21] The Withdrawal Agreement and the Northern Ireland Protocol were adopted at the end of 2019. They regulate the terms of the UK's withdrawal from the EU. The Protocol is a special appendix to the Withdrawal Agreement regarding Northern Ireland's status after Brexit. The TAC was agreed in December 2020. It shall govern the future relations of the UK as a non-member state with the EU. One of the key provisions of the Protocol is Article 5, which outlines the role for the joint committee. The joint committee consists of representatives of the EU and the UK who, in the event of any issues occurring with the functioning of the Protocol, shall meet and attempt to find an acceptable solution.

From the UK's point of view, there are four principal statutes relating to Brexit: the European Union (Withdrawal) Act 2018,[22] which retains EU law in the UK legal system so as to allow legal continuity, the European Union (Withdrawal Agreement) Act 2020,[23] which gives effect in UK law to the revised legal agreement, the United Kingdom Internal Market Act 2020,[24] which concerns trade between the different nations of the UK, necessitated by the Northern Ireland Protocol, which is the focus of this article, and the European Union (Future Relationship) Act 2020,[25] which implements the December 2020 EU-UK Trade and Cooperation Agreement[26] into the UK domestic legal order. These pieces of British legislation shall be mentioned later in the section regarding the role of EU law under the Protocol. The British government flirted with breaking

18 ibid.

19 Withdrawal Agreement (n 8).

20 Northern Ireland Protocol (n 2).

21 Trade and Cooperation Agreement between the European Union and the European Atomic Energy Community, of the one part, and the United Kingdom of Great Britain and Northern Ireland, of the other part (European Commission, 24 December 2020) <https://ec.europa.eu/info/sites/info/files/draft_eu-uk_trade_and_cooperation_agreement.pdf> accessed 18 April 2021.

22 European Union (Withdrawal) Act 2018 c.16.

23 European Union (Withdrawal Agreement) Act 2020 c.1.

24 United Kingdom Internal Market Act 2020 c.27.

25 European Union (Future Relationship) Act 2020 c.29.

26 Trade and Cooperation Agreement Between the European Union and the European Atomic Energy Community, of the one part, and the United Kingdom of Great Britain and Northern Ireland, of the other part [2020] OJ L444.

international law in introducing the Northern Ireland Protocol Bill in Parliament. As a consequence of the actions of the British government, trust between the EU and the UK is currently low.

II. The Impact of Brexit and the Protocol on the Economy of Northern Ireland

A. Northern Ireland's *de facto* membership of the single market and customs union

The document adopted by the EU and the UK regarding the future of Northern Ireland and Ireland's relationship with each other and the EU is titled "The Protocol on Ireland-Northern Ireland." It is attached to the withdrawal agreement. In order to achieve the aims agreed upon regarding the island of Ireland, the parties to the negotiations agreed that Northern Ireland shall *de facto* remain attached to the EU's customs union and its internal market, and committed to its rules and institutions, while the remaining territory of the United Kingdom shall leave these institutions, and EU law shall have no effect in that territory, beyond the date of Brexit. The EU did not insist on such an outcome because of pure altruism. It had an interest in safeguarding its fiscal interests and regulations. Had the border between Northern Ireland and Ireland remained "soft," in the sense that there were no customs or border posts, without measures safeguarding the integrity of the EU's institutions and regulations, there was a risk that goods and persons could have entered the EU without the necessary checks being made. The government of Ireland sat at the table as a negotiator on the side of the EU and had an interest in avoiding a hard border and preserving the peace process.[27]

The first problem with the Protocol is that it is extremely complicated. While Northern Ireland remains *de facto* within the EU single market and customs union and committed to some, but not all, of the EU's laws, the island of Great Britain does not.[28] The Protocol has created uncertainty with regards to Northern Ireland's status. It remains "attached to the EU's customs union but with additions, while it is associated with the EU's internal market but with subtractions."[29] Article 4 of the Protocol states in plain English: "Northern Ireland is part of the customs territory of the United Kingdom." Simple enough, it is clear that Northern Ireland would be part of any future trade agreement the UK government chose to

[27] Lisa O'Carroll, 'Leo Varadkar: Brexit has undermined Good Friday agreement' *The Guardian* (London, 3 November 2018) <https://www.theguardian.com/world/2018/nov/03/leo-varadkar-brexit-has-undermined-the-good-friday-agreement> accessed 27 April 2021.

[28] Northern Ireland Protocol (n 2).

[29] ibid.

conclude, though other aspects of the Protocol complicate things.[30] This is an issue as Northern Ireland is subject to the EU law regime; this could potentially be complicated by the UK as a whole entering into a trade agreement which would contradict the Protocol.

An example of the Protocol complicating things would be the rules regarding customs duties. Article 5(1) of the Protocol provides that customs duties shall only be payable on goods moving from Great Britain into Northern Ireland if that good is at risk of being moved into the European Union, "whether by itself or forming part of another good following processing." Goods arriving from a third country, such as Canada, would also be subject to an EU tariff if they were "at risk".[31] This all seems innocuous. The problem is that the burden of proof will be on the importer or trader, who must prove (a) that the goods in question will not be subject to commercial processing in Northern Ireland; and (b) that they fulfil the criteria established by the joint committee in accordance with the fourth subparagraph. These are: (a) the final destination and use of the good; (b) the nature and value of the good; (c) the nature of the movement; and (d) the incentive for undeclared onward movement into the EU, in particular incentives resulting from the duties payable pursuant to paragraph 1. The definition of "commercial processing" is overly broad. "Commercial processing" is considered to be "any alteration […] or transformation of goods in any way." Even if commercial processing and end consumption take place entirely within Northern Ireland, there is still the likelihood of having to pay duties. An example would be of a widget maker in Belfast who imports components from an English supplier. This widget maker would now be liable to pay EU customs duties, as by using those components to make widgets he will subject them to a form of processing—it does not matter if the product actually enters the EU or not.[32] The definition in (b) has still to be defined and should no definition arise, it will be assumed that all goods in Northern Ireland are at risk of entering the EU. This seems like overreach. This rule preserves the integrity of the EU's internal market at the expense of the UK's.[33]

Moreover, Great Britain accounted for 60 per cent of all goods imported into Northern Ireland in 2018, more than four times the amount imported from Ireland.[34] While preserving the Northern Ireland peace process was one of the

30 Alfred Artley and George Peretz, 'Customs and the Northern Ireland Protocol' (Monckton Chambers, 17 April 2020) < https://www.monckton.com/wp-content/uploads/2020/04/TJ_2020_Issue1483_Apr_Peretz-002.pdf> accessed 13 November 2020.

31 ibid.

32 Artley and Peretz (n 30).

33 Stephen Weatherill, 'The Protocol on Ireland/Northern Ireland: protecting the EU's internal market at the expense of the UK's' (2020) 45(2) European Law Review 222.

34 Padraic Halpin and Kate Holton, 'Northern Ireland looks south as Brexit takes bite out of UK trade links' (*Reuters*, 23 December 2020) <https://www.reuters.com/article/us-britain-eu-nireland/northern-ireland-looks-south-as-brexit-takes-bite-out-of-uk-trade-links-idUSKBN28X0Q3> accessed 11 January 2021.

EU's stated aims with regards to the Protocol, what the Protocol could do to Northern Ireland's economy is concerning. By hampering its ability to trade with its largest partner, Great Britain, there runs the risk of placing an already deprived region into further economic pressure. While the decision to ensure the border remained as it was prior to Brexit deserves praise, it does seem that the interests of the nationalist population[35] were placed above those of the unionists, who favour close political and economic ties with Great Britain. It was not the EU's fault that the British government was ideologically committed to withdrawing from the EU as well as the single market and customs union. Had the entire United Kingdom remained in the customs union and single market, the problem with regards to Northern Ireland being unable to trade with its largest market would not have arisen. It is unfortunate that both the UK and the EU were so committed to their targets—the UK leaving the EU entirely and the EU preserving the integrity of its single market—that it became impossible to reconcile the red lines of the UK and the EU in the negotiations, and the interests of Northern Ireland were undermined.

Article 5(3) of the Protocol reveals that the entirety of EU customs law shall apply in Northern Ireland.[36] Although the Preamble to the Protocol,[37] Article 4 of the Protocol,[38] and Prime Minister Johnson all claim differently,[39] Northern Ireland is *de facto* part of the EU's customs territory.

B. The Article 16 Safeguard Mechanism and its Potential to Impact the Economy of Northern Ireland

Article 16(1) of the Northern Ireland Protocol provides: "If the application of this Protocol leads to serious economic, societal, or environmental difficulties that are liable to persist, or to diversion of trade, the Union or the United Kingdom may unilaterally take appropriate safeguard measures, Such safeguard measures shall be restricted with regard to their scope and duration to what is strictly

[35] Namely to prevent a hard border and maintain the all-Ireland economy.

[36] Artley and Peretz (n 30).

[37] The Preamble states: 'Noting that nothing in this Protocol prevents the United Kingdom from ensuring unfettered market access for goods moving from Northern Ireland to the rest of the United Kingdom's internal market;' and 'Recalling that Northern Ireland is part of the customs territory of the United Kingdom and will benefit from participation in the United Kingdom's independent trade policy.'

[38] Article 4 of the Protocol provides: 'Northern Ireland is part of the customs territory of the United Kingdom.'

[39] See Patrick Daly and Megan Baynes, 'Johnson tells Northern Ireland businesses to 'bin' customs forms' *Belfast Telegraph* (Belfast, 8 November 2019) <https://www.belfasttelegraph.co.uk/news/northern-ireland/johnson-tells-northern-ireland-businesses-to-bin-customs-forms-38674258.html> accessed 18 April 2021. Per the article: "Mr Johnson clarified further when he told reporters: 'Northern Ireland and the rest of GB are part of the UK customs territory and there can be no checks between goods operating in one customs territory. We're the UK. We will not be instituting such checks."

necessary in order to remedy the situation. Priority shall be given to such measures as will least disturb the functioning of this Protocol."

Article 16 is a "last resort" provision.[40] It exists to allow either the EU or the UK to take unilateral action in response to negative effects arising from the Protocol. The safeguard allows the UK and the EU to take unilateral action if the Protocol is leading to "economic, societal, or environmental difficulties."[41] Another circumstance in which Article 16 may be invoked is when a "diversion of trade" occurs.[42] A problem with this provision for both parties to the Agreement is that this wording is vague and does not provide a clear example of when invoking Article 16 would be invoked under those criteria.

Article 16 is not a route to the unilateral disapplication of the Protocol.[43] Nor is it a "route" to unilateral suspension.[44] In the event it was triggered, the Protocol would continue to apply, and so would the obligations that derive from it. The process to be followed upon triggering Article 16 is as follows; if either party is considering adopting safeguard measures unilaterally, it must notify the other party "without delay" and through the joint committee. The party must provide all "relevant information," details of why unilateral action is needed, what the proposed action is, and justification for it.[45] There is then supposed to be a consultation period where the two parties work out a mutually acceptable solution. If such a unilateral safeguard is adopted, the joint committee must be made aware of it and discuss them within three months with a view to abolishing it as soon as possible. None of this occurred during the brief period Article 16 was invoked by the EU.[46]

In early 2021, the EU made an "aborted" attempt[47] to trigger this safeguard which was reversed within hours after condemnation from the UK.[48] The decision to trigger Article 16 was made in response to fears that Northern Ireland could be used as a "back door" to get around restrictions and send more supplies of the vaccine to Great Britain.[49]

[40] Katy Hayward and David Phinnemore, 'Article 16 of the Ireland/Northern Ireland Protocol offers no 'quick fix' (*London School of Economics*, 14 January 2021) <https://blogs.lse.ac.uk/europpblog/2021/01/14/article-16-of-the-ireland-northern-ireland-protocol-offers-no-quick-fix/> accessed 15 February 2021.

[41] ibid.

[42] ibid.

[43] Hayward and Phinnemore (n 40).

[44] ibid.

[45] ibid.

[46] Lisa O'Carroll, 'EU's article 16 blunder has focused minds on Northern Ireland' *The Guardian* (London, 4 February 2021) <https://www.theguardian.com/uk-news/2021/feb/04/eus-article-16-blunder-should-focus-minds-on-northern-ireland> accessed 15 February 2021.

[47] ibid.

[48] 'What is Article 16 and why did the EU make a U-turn after triggering it?' (*Sky News*, 31 January 2021) <https://news.sky.com/story/what-is-article-16-and-why-did-the-eu-make-a-u-turn-after-triggering-it-12202915> accessed 15 February 2021.

[49] ibid.

The EU's aborted attempt to trigger Article 16 has led to more ambiguity surrounding this provision, as the UK government has seized upon the subsequent controversy as a means of demanding an extension to the post Brexit "grace period".[50] A shortage of AstraZeneca COVID-19 vaccine doses would surely constitute an "economic, social or environmental difficult(y)" or perhaps a "diversion of trade" given the European Commission's suspicions that vaccines were moving from the single market into Great Britain through Northern Ireland, though the problem was that the Commission chose to act on suspicions without any apparent solid evidence.[51]

Relations between the EU and the UK were strained in late 2020 by the apparent willingness of the UK government to breach international law and renege on its commitments under the Withdrawal Agreement (during the period in which the Internal Market Bill was passing through Parliament).[52] The strains were further exacerbated by the EU's decision to trigger Article 16 in January 2021. Overall, the potential for either side to trigger Article 16 is a very real possibility and a problem which must be overcome in order to make the Protocol work.

C. Possible solutions to customs and goods disputes

The next section will focus on the potential solutions to the issue of customs and goods disputes under the Protocol. Two solutions are offered: one UK wide; the other Northern Ireland only. The latter would be more feasible, as the current British government has expressed its distaste for the entirety of the UK remaining regulatorily aligned with the EU.

(i) Solutions to Customs and Goods Disputes: A Potential UK Wide Solution

There is still uncertainty in some areas with regards to how the Protocol will function.[53] However, it is clear from recent events that the EU-UK joint committee and the Court of Justice of the European Union (CJEU) will play a role in resolving disputes and giving clarity to ambiguous provisions of the Protocol. The Protocol has led to checks and controls being imposed on goods moving from

50 O'Carroll (n 46).

51 Daniel Boffey and Kim Willsher, 'EU in U-turn over move to control vaccine exports to Northern Ireland' *The Guardian* (London, 29 January 2021) <https://www.theguardian.com/world/2021/jan/29/eu-controls-on-vaccine-exports-to-northern-ireland-trigger-diplomatic-row> accessed 15 February 2021.

52 Hayward and Phinnemore (n 40).

53 Brendan McGurk, 'Analysis of the Northern Ireland Protocol and its impact on the UK' (*LexisNexis*, 3 July 2020) <https://www.lexisnexis.com/uk/lexispsl/lifesciences/document/412012/608H-BYT3-GXFD-809N-00000-00?utm_source=psl_da_mkt&utm_medium=referral&utm_campaign=analysis-of-the-northern-ireland-protocol-and-its-impact-on-the-uk> accessed 15 April 2021.

Great Britain to Northern Ireland.[54] The problems the Protocol has had for businesses led to the UK government unilaterally extending grace periods[55] for food products moving from Great Britain to Northern Ireland—an example of the UK violating the Withdrawal Agreement, and therefore international law.[56] There is the prospect of non-legal action the EU could take against the UK for failing to honour its commitments. Some EU diplomats have suggested retaliation against the UK through the financial services industry. Presumably this means restricting the access of British financial firms to the EU financial market.[57]

It has already been suggested by commentators that the Protocol does not amount to a permanent "fix" with regards to the issue of goods moving into and out of Northern Ireland.[58] One example is that of food products. The requirement for Sanitary and Phytosanitary (SPS) checks has disrupted the movement of food products from Great Britain to Northern Ireland. The time required for these checks has had an impact on the food supply in Northern Ireland, with bare shelves in supermarkets now being a common sight. The problems this is causing would suggest the need for a new solution, yet the two existing arrangements the EU has with Switzerland and New Zealand do not seem to be viable as a solution for Northern Ireland.[59] The agreement Switzerland has with the EU requires Switzerland to adopt all relevant EU legislation to prevent the need for checks. Because the UK has ruled out indefinitely aligning with EU regulations, this model will not be adopted. The New Zealand model would not require alignment, but it would acknowledge each party's SPS standards and reduce the percentage of checks required. The EU and the UK should already have reached such an agreement.[60] Yet, such an arrangement would not solve Northern Ireland's problems, for the issue is the kinds of checks required, not the amount. A new solution is required.

There are two possible solutions: one UK wide, the other specific to Northern Ireland. The UK wide model would require the UK and the EU to conclude a new SPS agreement that would manage the divergence of regulations and limit the need for checks. A precedent in managing diverging standards between the two parties was struck in 2020 by the Trade and Cooperation

54 Raoul Ruparel, 'How to fix Brexit's Northern Ireland protocol problem' (*Politico*, 26 March 2021) <https://www.politico.eu/article/brexit-northern-ireland-protocol-border-checks-eu-uk-agreement/> accessed 17 April 2021.

55 In the context of law, a grace period is a time period during which a particular rule exceptionally does not apply, or only partially applies.

56 Jacopo Barigazzi and Hans von der Burchard, 'EU countries back legal action against UK over post-Brexit grace period extension' (*Politico*, 9 March 2021) <https://www.politico.eu/article/eu-countries-back-legal-action-against-uk-over-post-brexit-grace-period-extension/> accessed 17 April 2021.

57 Ruparel (n 54).

58 ibid.

59 ibid.

60 ibid.

Agreement (TCA).[61] In the TCA, the UK agreed not to regress[62] its labour and social employment laws.[63] If the UK were to fail to meet its commitments, the EU would be able to take remedial measures, which is enforced and overseen by a panel of experts.[64] A rebalancing clause is also included,[65] which in the event of the standards of the UK and the EU diverging, would allow either side to take measures[66] to manage the divergence of standards. These processes provide a mechanism for managing divergence which can be used elsewhere.[67] While the mechanisms in the TCA are focussed on avoiding tariffs on goods, they could potentially be applied in the context of checks on goods as well. By transposing these mechanisms to the context of goods, a solution which could potentially fix the problems Northern Ireland is currently going through is available.

This model would be similar to the Swiss model in the particular area of food regulations, but it would potentially be more robust and offer more assurance to the EU, as it has a resolution mechanism in place already, which the Swiss model does not. It also would not require any changes to EU law. The UK would be meeting the EU's legal requirements while gaining the benefits of the agreement, which would be withdrawn if it broke the hypothetical agreement.[68]

(ii) A Northern Ireland Only Solution

A UK-wide solution would not be politically workable, as the current UK government wishes to remain unaligned with EU standards.[69] There is an option which would apply to Northern Ireland only. In section B, the concept of goods being "at risk" of being sold in the EU was mentioned. A solution for the issue of agri-foods would be to extend the category of "at risk" goods to include agri-food regulations. If there is no risk of food products being sold in the EU, then they

61 The EU-UK Trade and Cooperation Agreement' (*European Commission*, 31 December 2020) <https://ec.europa.eu/info/relations-united-kingdom/eu-uk-trade-and-cooperation-agreement_en> accessed 17 April 2021.

62 To ensure that its labour and social employment laws did not fall behind EU standards.

63 Trade and Cooperation Agreement between the European Union and the European Atomic Energy Community, of the one part, and the United Kingdom of Great Britain and Northern Ireland, of the other part [2020] OJ L-444, chapter six: labour and social standards, article 6.2(1)-(2).

64 ibid, article 6.4(1)–(2).

65 ibid, article 9.4.

66 'Rebalancing measures' is a rather broad term, but it essentially means that if the actions of one-party lead to 'material impacts', in this case on labour and social protection, the other side may take proportionate action to restore the balance. An arbitration tribunal has been set up to mediate in potential disputes. See David Glass, 'Brexit update: ESG reporting, rebalancing measures and trade with India' (*Excello Law*, 22 February 2021) <https://excellolaw.co.uk/excellolaw-blogs/brexit-update-esg-reporting-rebalancing-measures-trade-with-india/> accessed 17 April 2021.

67 Ruparel (n 54).

68 ibid.

69 Kenneth Armstrong, 'Regulatory autonomy after EU membership: alignment, divergence and the discipline of law' (2020) 45(2) European Law Review 207.

would be exempt from agri-food requirements. This argument is also predicated on the UK and the EU agreeing to extend the Brexit grace period, something which does not seem likely. It would also depend on the use of data to allow firms to prove their goods are only sold in Northern Ireland. This would require some derogation from EU law, but the Northern Ireland Protocol already does this by keeping Northern Ireland in the single market for goods and the customs union but not requiring the other fundamental freedoms. The EU would be wise to address this issue, as would the UK. Both sides agreed that Brexit would "impact as little as possible on the everyday life of communities in both Ireland and Northern Ireland."[70]

The problem with this argument is that the category of "at risk" goods is very wide.[71] A good is seen as being "at risk" of moving into the EU unless it can be proven that it will not be subject to commercial processing in Northern Ireland and that it meets the criteria drawn up by the joint committee established by Article 164 of the Withdrawal Agreement.[72] Article 5(2) of the Protocol defines "commercial processing" as any alteration or transformation of goods. An example would be flour imported into Northern Ireland from Great Britain. This flour would be subject to EU customs even if the bread made from it was not intended to be sold outside of Belfast.[73] Further, the wording of Article 5(2) reflects that the default position is that duties will have to be paid on goods moving from Great Britain to Northern Ireland, unless it can be proven that the good is not "at risk". The broad definition of "commercial processing" means[74] that a Northern Ireland only solution would not be a feasible solution. Overall, the UK wide model would be the better solution for the issue of goods moving between Great Britain and Northern Ireland.

D. ARE THE CURRENT ARRANGEMENTS PROPORTIONATE?

A key concept in the law of the EU is the idea of proportionality: whether a measure taken by a member state which derogated from the rights conferred by the Treaties was "not… beyond that which is necessary in order to achieve the objective. In other words, it must not be possible to obtain the same result by less restrictive rules."[75] The same question might be asked of certain aspects of the Protocol, particularly with regard to its rules on the movement of goods between Britain and Northern Ireland. The EU must consider whether the current

[70] Preamble to the Northern Ireland Protocol.
[71] Artley and Peretz (n 30).
[72] Article 5(2) Northern Ireland Protocol.
[73] Weatherill (n 33).
[74] Artley and Peretz (n 30).
[75] Case C-288/89 *Gouda* [1991] ECR I-4007, para 15.

arrangements are proportionate to the risk posed to the single market.[76] In January 2021, 36,000 point-of-entry certificates were required across the entire EU. Of this total, 5,800 were required for trade between Britain and Northern Ireland, which represents 15 per cent, even though trade between Northern Ireland and Britain constitutes less than 1 per cent of total EU trade with non-EU countries.[77] Some more statistics showing the scale of the problem: Northern Ireland is processing more paperwork than any EU member state for animal imports; is processing 20 per cent of all CHED-Ps[78] in the EU; and up to 90 per cent of generic drugs could be withdrawn from Northern Ireland because medicines made in Great Britain have to be licensed separately for use in the region as well as undergo separate checks.[79]

This situation has arisen because of the lack of trust between the EU and the UK because of the UK's actions (the UK's strategy to secure concessions from the EU has been to be antagonistic towards the EU).[80] The Protocol is "the only show in town", at least while the UK is led by hard-line Brexiters, to protect the EU's vital interests, but in the future, the EU and the UK will have to consider whether the Protocol is *too* restrictive on trade between Great Britain and Northern Ireland. Since the agri-food checks are proving most problematic, perhaps that will be the first area for reform.

III. The Legal Issues Arising from the Protocol

A. The continuing role of EU law and the CJEU in Northern Ireland

Article 12 of the Protocol mandates that the United Kingdom is responsible for ensuring the application of relevant EU law.[81] This, of course, is a risk for the EU,

76 Jess Sargeant, 'The UK government must take responsibility for making the Northern Ireland protocol work' (*Institute for Government*, 15 April 2021) <https://www.instituteforgovernment.org.uk/blog/northern-ireland-protocol-tensions> accessed 19 April 2021.

77 Sam McBride, 'Unionist leaders unite to go to court over Irish Sea border, arguing it breaches the 1800 Act of Union and 1998 Belfast Agreement' *News Letter* (Belfast, 21 February 2021) <https://www.newsletter.co.uk/news/politics/unionist-leaders-unite-to-go-to-court-over-irish-sea-border-arguing-it-breaches-the-1800-act-of-union-and-1998-belfast-agreement-3141841> accessed 19 April 2021.

78 A form for importing animal products.

79 Noelle McElhatton, 'EU Brexit chief says trade friction in Northern Ireland can be sorted but adds the task is 'massive' (*Export*, 19 April 2021) <https://www.export.org.uk/news/561278/EU-Brexit-chief-says-trade-friction-in-Northern-Ireland-can-be-sorted-but-adds-the-task-is-massive.htm> accessed 19 April 2021.

80 Maddy Thimont Jack, 'Making Lord Frost cabinet minister for EU relations makes sense – and suggests a hostile strategy' (*Institute for Government*, 18 February 2021) <https://www.instituteforgovernment.org.uk/blog/lord-frost-brexit-cabinet-minister> accessed 19 April 2021.

81 Oliver Garner, 'The new Irish Protocol could lead to the indefinite jurisdiction of the EU Court of Justice within the UK' (*London School of Economics*, 23 October 2019) <https://blogs.lse.ac.uk/brexit/2019/10/23/the-new-irish-protocol-could-lead-to-the-indefinite-jurisdiction-of-the-court-of-justice-of-the-european-union-within-the-united-kingdom/> accessed 15 April 2021.

as it means that the EU has outsourced the patrolling of its external border on the island of Ireland to the UK, a third country.[82] The remainder of Article 12, however, ensures the EU institutions are able to supervise the UK's application of EU law. Articles 12(4) and 12(5) accord the CJEU jurisdiction over the application of key provisions of the Protocol.[83] EU actors, including the CJEU, shall retain the powers and jurisdiction accorded to them by the Treaties[84] in this regard. The Article 267 TFEU preliminary reference procedure shall continue to apply to and in the United Kingdom. This means that individuals who are prevented from benefiting from the UK's enforcement of EU law will have the ability to bring a claim before a domestic court in the United Kingdom, and that court will be required to refer the issue to the CJEU if the relevant criteria are fulfilled.[85]

A preliminary ruling is a ruling by the CJEU on "(a) the interpretation of the Treaties; (b) the validity and interpretation of acts of the institutions, bodies, offices or agencies of the Union."[86] If, as in the case of *Factortame*,[87] the CJEU finds that a member state's[88] legislation conflicts with EU law, the member state will be required to 'disapply' such law, though the CJEU does not itself have the power to amend such law. The *Francovich*[89] principle of state liability may continue to apply in the event that individuals were prevented from benefitting from EU law by the UK's failure to enforce it. Such individuals would have the ability to petition a UK domestic court, and the domestic court would then have to submit the question to the CJEU, pending the criteria being fulfilled.[90]

Article 4 of the withdrawal agreement provided for the continuing jurisdiction of the CJEU in Great Britain until the end of the transition period[91] on 31 December 2020. The CJEU now only has jurisdiction on the issue of the rights of EU citizens in Great Britain.[92] In Northern Ireland, however, the Protocol gives the CJEU the ability to potentially rule upon the actions of UK authorities indefinitely. This power could only be terminated by the Northern Ireland Assembly refusing to consent to its continuing operation beyond 2024 (which is within the realm of possibility given the current political situation in Northern

82 ibid.

83 Craig and de Búrca (n 6) 779.

84 Meaning the Treaty on European Union (TEU) and the Treaty on the Functioning of the European Union (TFEU).

85 Garner (n 81).

86 Article 267 TFEU.

87 Case C-213/89 *R v Secretary of State for Transport, ex parte Factortame Ltd and others* [1990] ECR I-2433, para 23.

88 The UK is no longer a member state but given that the Northern Ireland Protocol requires UK courts to ask for a preliminary ruling in cases involving the application of EU law in Northern Ireland, in the context of the Northern Ireland Protocol, the requirement for the arbiter to be a court of a member state is null.

89 Cases C-6/90 and C-9/90 *Francovich and Bonifaci and others v Italy* [1991] ECR I-5357.

90 Garner (n 81).

91 The transition period was a period when the entirety of the UK (not just Northern Ireland) remained in the EU customs union and single market and followed EU rules.

92 Garner (n 81).

Ireland) or by a future agreement which would supersede the Protocol and the CJEU.[93]

Differing from the prior withdrawal agreement which Theresa May's government concluded with the EU, the final withdrawal agreement and the attached Northern Ireland Protocol provide for a heightened role for the CJEU (most likely as a result of only Northern Ireland remaining *de facto* attached to the EU's Customs Union and single market rather than the entire UK). The previous withdrawal agreement made provision for a role for the CJEU, however, this would have been mitigated by the EU and UK "engag[ing] in best endeavours"[94] to prevent the Northern Ireland backstop[95] coming into force. The CJEU would have had a role only in the event the backstop had to be enforced. The Protocol, however, mandates that the CJEU will have a role for as long as the Protocol remains in force.

It is more likely that litigation will arise under the current agreement[96] than it would have under the one Theresa May's government negotiated;[97] this is because whereas under the previous withdrawal agreement, the CJEU would have only been determining whether the UK was complying with customs union rules that the UK had already adopted during membership, under the current withdrawal agreement, the CJEU will have to ensure the UK checks goods movements using the criteria established by the joint committee. The fact that these criteria are completely new to both individuals and state actors means that it is likely the CJEU will need to at some point give clarity to questions of law. Furthermore, if the UK were to adopt different regulatory standards from the EU, there would be questions regarding the ability of its authorities to enforce EU law standards in Northern Ireland as well.[98]

There have already been several occasions where the UK government has either flirted with disregarding its commitments under the Withdrawal Agreement and the Northern Ireland Protocol,[99] or shown a lack of trustworthiness in its

93 Article 13(8) Northern Ireland Protocol.

94 Garner (n 81).

95 The 'backstop' was the solution to the Irish border issue that was negotiated by Theresa May's government with the EU. It was replaced by the Northern Ireland Protocol. The backstop would have kept Northern Ireland in some parts of the EU single market, until the EU and the UK agreed on a long-term solution. The EU and UK customs territories would have operated as one until a long-term arrangement was agreed upon. See Jon Henley, 'Brexit deal: key points from the draft withdrawal agreement' *The Guardian* (London, 14 November 2018) <https://www.theguardian.com/politics/2018/nov/14/brexit-deal-key-points-from-the-draft-withdrawal-agreement> accessed 16 April 2021.

96 Withdrawal Agreement (n 8).

97 'Progress on the UK's exit from, and future relationship with, the European Union' (*Department for Exiting the European Union*, 14 November 2018) <https://www.gov.uk/government/publications/progress-on-the-uks-exit-from-and-future-relationship-with-the-european-union> accessed 16 April 2021.

98 Garner (n 81).

99 Oliver Garner, 'A Barrier against the new incoming tide? The UK Internal Market Bill and Dispute Resolution under the Withdrawal Agreement and the Protocol on Ireland/Northern Ireland' (*UK Constitutional Law Association*, 17 September 2020) <https://ukconstitutionallaw.org/2020/09/17/oliver-garner-a-barrier-against-the-new-

conduct during the Brexit process. In late 2019, the UK government considered not requesting the European Council to extend the Article 50 negotiations period, despite domestic UK law requiring it to do so.[100] Exactly one year later, Northern Ireland Secretary Brandon Lewis admitted that the UK Internal Market Bill which was passing through the UK Parliament at the time would "break international law" and go against the Withdrawal Agreement in a "specific and limited way".[101] The Bill would have given government ministers the power to define what state aid needs to be reported to the EU[102] and products that are at risk of being brought into Ireland from Northern Ireland.[103]

The plans of the British government were dropped after the EU and UK were able to come to agreement through the joint committee,[104] but what would happen if the British government were to breach provisions of the Northern Ireland Protocol and ignore the protests of the EU? Article 12 of the Protocol is the starting place.

B. HOW WILL THE EU ENSURE EU LAW IS ENFORCED IN NORTHERN IRELAND?

It seems the most likely legal solution to disputes regarding EU law in Northern Ireland for the moment will be for the enforcement mechanisms available under EU law to be utilised.[105] Per Article 12 of the Protocol, there would be two mechanisms: the first would allow the European Commission to bring an infringement claim against the UK before the CJEU; the second would allow an individual in the UK to bring a case before a national court. The court would then possibly have to initiate a preliminary reference procedure.[106] The extent to which the second mechanism would operate in UK law is unclear. The British government has already shown a willingness to disregard its obligations under the Protocol. In a legal opinion, the government held that "Parliament's ability to pass

incoming-tide-the-uk-internal-market-bill-and-dispute-resolution-under-the-withdrawal-agreement-and-the-protocol-on-ireland-northern-ireland/> accessed 16 April 2021.

100 'Brexit extension: PM to 'test law to limit' to avoid delay' (*BBC News*, 8 September 2019) <https://www.bbc.co.uk/news/uk-politics-49625431> accessed 16 April 2021.

101 'Northern Ireland Secretary admits new bill will 'break international law' (*BBC News*, 8 September 2020) <https://www.bbc.co.uk/news/uk-politics-54073836> accessed 16 April 2021.

102 EU level playing field provisions continue to apply to Northern Ireland, per Article 10(1)-(3) of the Northern Ireland Protocol.

103 Daniel Boffey, Jessica Elgot and Heather Stewart, 'Leaked EU cables reveal growing mistrust of UK in Brexit talks' *The Guardian* (London, 7 September 2020) <https://www.theguardian.com/politics/2020/sep/07/leaked-eu-cables-reveal-mistrust-of-uk-motives-in-brexit-talks> accessed 16 April 2021.

104 Emer O'Toole, 'Brexit: Tories remove law-breaking clauses from Internal Market Bill in U-turn' *The National* (Glasgow, 8 December 2020) <https://www.thenational.scot/news/18929161.brexit-tories-remove-law-breaking-clauses-internal-market-bill-u-turn/> accessed 16 April 2021.

105 Garner (n 99).

106 ibid.

provisions that would take precedence over the Withdrawal Agreement was expressly confirmed in section 38 of the European Union (Withdrawal Agreement) Act 2020, with specific reference to the EU law concept of 'direct effect'."[107] The British government wished to eliminate the possibility of relying on the direct effect of the withdrawal agreement provisions.

The UK's having a dualist[108] legal system raises the possibility of the preliminary reference procedure being suspended.[109] The second paragraph of the legal opinion states clearly: "Clause 45 of the Bill partially disapplies the implementation in UK domestic law of Article 4 WA and the EU law concept of direct effect."[110] How far this "partial" disapplication of direct effect was intended to go is unclear. The objective of this provision is arguably to prevent domestic courts from hearing challenges to the legislation and issuing preliminary references under Article 267 TFEU.[111] Indeed, the very concept behind direct effect is that the EU Treaties create legal rights which can be enforced by both natural and legal persons before the courts of the EU's member states.[112] There is also the interesting theoretical question regarding whether these rights would be directly effective in a *former* member state. Per Article 12(7)(a) of the Protocol: "The United Kingdom may participate in proceedings before the Court of Justice of the European Union in the same way as a member state" with regards to issues arising from the Protocol. So the answer would be *de jure* "yes"; though how this would play out in practice is up for debate.

C. What if the British Government disregards the application of EU law?

Given that the withdrawal agreement requires the direct effect of EU law in Northern Ireland and a role for the CJEU, the UK no longer being a member state is irrelevant, though a potential case in the CJEU to provide more clarity would be helpful. The mere fact a country is a member state of the EU does not in itself guarantee compliance with the judgments of the CJEU. In the *Ajos*[113] case,

107 'HMG Legal Position: UKIM Bill and Northern Ireland Protocol' (*Cabinet Office*, 10 September 2020) <https://www.gov.uk/government/publications/hmg-legal-position-ukim-bill-and-northern-ireland-protocol> accessed 17 April 2021.

108 In international law, each state can choose the relationship between domestic law and international law within its legal system. Two theories exist. A 'monist' legal system incorporates international law into the domestic legal order. International law would apply as though it were domestic law. In a 'dualist' legal system, international law is seen as the law between states, national law is the law within a state. While international law is binding at the international level, it cannot be binding within the domestic legal system. See Robert Schütze, *European Union Law* (2nd edn, Cambridge University Press 2018) 77.

109 Garner (n 99).

110 HMG Legal Position (n 107).

111 Garner (n 99).

112 Case 26/62 *Van Gend en Loos v Nederlandse Administratie der Belastingen* [1963] ECR 13.

113 Case C-441/14 *Dansk Industri v Rasmussen* [2016] EU:C:2016:278.

the Danish Supreme Court initiated an Article 267 preliminary reference procedure regarding the compatibility of paragraph 2(a)(3) of the Danish Salaried Employees Act with Directive 2000/78/EC.[114] Despite the clear guidance given by the CJEU, the Danish Supreme Court refused to set aside the provision of national law which was incompatible with the Directive, applying the national law instead.[115] The Danish court chose to follow this reasoning because it felt that disapplying the domestic legislation in favour of the EU Directive was *ultra vires*,[116] as under the Danish constitutional order, the judiciary should not issue a ruling which goes against the intention of the Danish parliament.[117]

This was a clear violation of the doctrine of primacy of EU law and loyal cooperation of national courts established by the *Costa* ruling.[118] National courts setting their own standards for the enforcement of EU law has been an ongoing development throughout the years. For instance, the German Constitutional Court set its own standards for enforcement of EU law in the case of *Internationale Handelsgesellschaft*,[119] where it held that it would enforce EU law "so long" as it complied with the fundamental rights guaranteed by the German legal order.[120] The point is with regards to the Protocol that courts of the member states have shown an unwillingness to enforce EU law over domestic law, so who is to say the courts of a *former* member state will not do the same?

Whether direct effect has been removed is now a moot issue, first because the UK government removed the relevant clauses from the Internal Market Bill.[121] The second reason being even if the UK government were to remove the direct effect of EU law and the ability of a UK court to bring an Article 267 TFEU preliminary reference proceeding, the European Commission still has the power to bring an infringement procedure against the UK before the CJEU as a matter of bilateral international treaty law.[122]

The act of introducing a bill into the UK Parliament with the offending clauses present was a breach of the international law maxim of *pacta sunt servanda*[123] and also a violation of Article 26 of the Vienna Convention on the Law of

114 Directive 2000/78/EC Framework Employment Equality [2000] OJ L303/16.

115 Case no. 15/2014 *Dansk Industri (DI) acting for Ajos A/S vs The estate left by A.*

116 Latin for "beyond the powers;" used in the sense that under Montesquieu's theory of the separation of powers, the judiciary should not play a legislative role by invalidating legislation and so on.

117 Mikael Rask Madsen, Henrik Palmer Olsen and Urška Šadl, 'Legal Disintegration? The Ruling of the Danish Supreme Court in AJOS' (*Verfassungsblog*, 30 January 2017) <https://verfassungsblog.de/legal-disintegration-the-ruling-of-the-danish-supreme-court-in-ajos/> accessed 18 April 2021.

118 Case 6/64 *Costa v. ENEL* [1964] ECR 585.

119 BVerfGE 37, 271 (*Solange I (Re Internationale Handelsgesellschaft)*) [1974] 2 CMLR 540.

120 Schütze (n 108) 130.

121 Which has since been passed as the United Kingdom Internal Market Act 2020 c.27.

122 Garner (n 99).

123 Latin for "agreements must be kept."

Treaties.[124][125] The outcome if in the future the UK government attempts to breach the Protocol in a similar way would be the commencement of a dispute resolution procedure in the joint committee, which would be established three months after written notification between the parties. In the event the joint committee was unable to find a solution, Article 12 of the Protocol would empower the European Commission to bring an infringement procedure against the UK before the CJEU.[126]

IV. The Political Problems Caused by Brexit, the Protocol's Attempted Solutions to Them, and Some Alternative Solutions

A. Was there a need for the border to be where it is?

The main problem during Brexit with regards to Northern Ireland was the clash of interests between the EU and the UK – there was never an optimal solution, and possibly never will be, to the extraordinarily complex problems brought about by the UK's withdrawal. There was never any solution that would have had no border on the island of Ireland, no border between Great Britain and Northern Ireland, and the ability for the entire UK to leave the EU single market and customs union.[127]

The phrase "to protect the Good Friday Agreement" may go down as one of those phrases used during Brexit which did not really mean anything, along with "Brexit means Brexit"[128] and "strong and stable".[129] While the Good Friday Agreement did see the withdrawal of British troops from the streets and fields of Northern Ireland, and did mean the border between Ireland and Northern Ireland became "invisible" in the sense there was no longer a military presence there, the border has been open since 1923 because of the Common Travel Area.[130]

124 Vienna Convention on the Law of Treaties (adopted 23 May 1969, entered into force 27 January 1980) 1155 UNTS 331.

125 Jonathan Deans, 'The Internal Market Bill: a specific and limited controversy?' (2021) 1 Juridical Review 48.

126 Garner (n 99).

127 Anton Spisak, 'After Brexit: Northern Ireland and the Future of the Protocol' (*Tony Blair Institute for Global Change*, 12 March 2021) <https://institute.global/policy/after-brexit-northern-ireland-and-future-protocol> accessed 19 April 2021.

128 Mark Mardell, 'What does 'Brexit means Brexit' mean?' (*BBC News*, 14 July 2016) <https://www.bbc.co.uk/news/uk-politics-36782922> accessed 19 April 2021.

129 Esther Addley and Caroline Davies, 'Dreadful night' when Theresa May's strong and stable fantasy evaporated' *The Guardian* (London, 9 June 2017) <https://www.theguardian.com/politics/2017/jun/09/theresa-mays-dreadful-night-strong-stable-fantasy-evaporated> accessed 19 April 2021.

130 'Memorandum of Understanding between the UK and Ireland on the CTA' (*Cabinet Office*, 8 May 2019) <https://www.gov.uk/government/publications/memorandum-of-understanding-between-the-uk-and-ireland-on-the-cta> accessed 19 April 2021.

As Rory Montgomery, former Irish diplomat, said:

> The Good Friday Agreement says either little or nothing about the European Union, about the border between North and South, or about trade within the UK. Therefore, the argument that Brexit or its outworkings formally violate the Agreement is hard to sustain. But Brexit seriously breaches the context and spirit of the Agreement, with very real political and psychological effects. One way or another, its implementation was always going to be disruptive and damaging.[131]

The solution to the trilemma did not necessarily have to be the one that was adopted in the end. The Good Friday Agreement was not violated by the act of Britain leaving the EU. The interpretation of the Good Friday Agreement which was presented by the government of Ireland became the widely accepted position and there was no attempt by the British government to articulate an alternative position.[132]

A "hard border" is not a legal term of art with a widely accepted definition that other terms such as *habeas corpus* or *ultra vires* have. It could have been interpreted differently. For instance, rather than have the border in the Irish Sea and thus keep Northern Ireland aligned to the single market and customs union, alternative arrangements could have been made to have checks on goods moving between the UK and Ireland some distance away from the actual border. If a land border between Ireland and Northern Ireland was having the economic and social impact the Protocol is having, it would be clear that it would need to be replaced.[133] While this may be true, looking at the empirical data makes it clear that more of Northern Ireland's trade is with Britain than with Ireland[134] and that the Protocol is impacting the supply of essentials like medicines.[135] From a utilitarian point of view, it would seem a border on the island of Ireland would have made more sense, although in the highly partisan politics of Northern Ireland, it would have been interpreted as favouring unionists over nationalists.

A border on the island of Ireland would have been more pleasing for the EU logistically, as it would have meant that it did not have to entrust its external

[131] Rory Montgomery, 'Protocol problems for both parts of Ireland: North and South' (*Fortnight Magazine*, April 2021) <https://fortnightmagazine.org/articles/protocol-problems-for-both-parts-of-ireland-north-and-south/> accessed 19 April 2021.

[132] ibid.

[133] Éilis O'Hanlon, 'Of course unionists are angry, we partitioned their country' *Irish Independent* (Dublin, 11 April 2021) <https://www.independent.ie/opinion/comment/of-course-unionists-are-angrywe-partitioned-their-country-40299253.html> accessed 19 April 2021.

[134] Halpin and Holton (n 34).

[135] McElhatton (n 79).

border to a third country, the UK. While the prospect of checks on goods could potentially have led to tensions within the nationalist community, the Good Friday Agreement, so often cited, does not provide that trade between Northern Ireland and Ireland should be unfettered. That could potentially have been solved in the future with some sort of alignment on SPS standards like the EU has with Switzerland and New Zealand, as Northern Ireland's trade with Ireland is primarily based on agricultural goods.[136]

The present situation with regards to the Irish border was not the only solution, but the result of a hard-line stance adopted by the then Taoiseach Leo Varadkar, which was accepted by the EU and the UK. Had the narrative reflected the reality that the Good Friday Agreement would not have been breached by checks on goods moving between Northern Ireland and Ireland, a solution more acceptable to all concerned parties may have been found.[137]

B. Article 18 and its Potential Lack of Compatibility with the Good Friday Agreement

Article 18 provides a mechanism for the UK to "provide the opportunity for democratic consent in Northern Ireland to the continued application of Articles 5 to 10."[138] This means that the UK government will have to seek the consent of the Northern Ireland Assembly, the local devolved legislature, in order for the Protocol's provisions to be extended beyond 2024.[139] Articles 5 to 10 cover, respectively, customs and movement of goods; protection of the UK internal market; technical regulations; VAT and excise; the single electricity market; and state aid.

Article 18 of the Protocol provides for "the opportunity for democratic consent in Northern Ireland […] consistent with the 1998 Agreement."[140] The 1998 Agreement is an international agreement between the UK and Ireland.[141] The "core tenet" of the Agreement, and the context of the reference to democratic consent in the Protocol, is that there should be no change to Northern Ireland's

136 Matthew Ward, 'Statistics on UK trade with Ireland' (*House of Commons Library*, 15 January 2021) <https://researchbriefings.files.parliament.uk/documents/CBP-8173/CBP-8173.pdf> accessed 19 April 2021.

137 Henry Hill, 'The Northern Ireland Protocol is untenable' (*The Spectator*, 16 April 2021) <https://www.spectator.co.uk/article/breaking-protocol-northern-ireland-s-brexit-deal-needs-reform> accessed 19 April 2021.

138 Northern Ireland Protocol, article 18.

139 Colin Harvey, 'Designing a Special Arrangement for Northern Ireland: the Irish Protocol in Context' (*Brexit Institute Working Paper Series*, 1 May 2020) <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3588415> accessed 16 February 2021.

140 The 1998 Agreement is the Good Friday or Belfast Agreement.

141 Austen Morgan, 'The Belfast Agreement: a practical legal analysis' (*CAIN Web Service*, 2000) <https://cain.ulster.ac.uk/events/peace/morgan/index.html> accessed 16 April 2021.

constitutional status without the consent of the unionist and nationalist communities.[142]

There is great dissatisfaction in Northern Ireland with the Protocol and its *de facto* creation of a border between Great Britain and Northern Ireland, particularly within the unionist community.[143] If there is a vote to bring the operation of Articles 5 to 10 to an end, then the joint committee will have two years to make recommendations to ensure the avoidance of a hard border.[144] It could even be the case that the joint committee will have to meet to discuss the abolition of the Irish Sea border. It would appear there is much for Brussels to be worried about: it would seem that the Northern Ireland Assembly would have the power to derail the Protocol, if it were so inclined. The role of the joint committee, however, acts as a constraint on the power of the Assembly. The CJEU will continue to have a role to play in matters of interpretation of EU law still in effect in Northern Ireland, and the joint committee will still be bound by the CJEU's rulings. Overall, this is a prickly provision of the Protocol, which has the potential to make politics in Northern Ireland even more heated, but it would seem the EU can rest easy knowing that its vital interest in securing its border with the UK will be protected by the might of the CJEU.

C. The potential incompatibility of the Protocol with the constitutional arrangements of Northern Ireland

The irony of the Protocol is that it was formulated by the EU and the UK with the intention of preventing further violence in Northern Ireland. In April 2021, the cities of Belfast and Derry-Londonderry had some of their worst riots in years. Loyalist paramilitaries have withdrawn their support for the Good Friday Agreement (which brought the most recent iteration of the conflict in Northern Ireland to an end) until the trading arrangements of the Northern Ireland Protocol are removed.

The EU showed a mature approach to the unrest by postponing the legal action it had intended to bring against the UK for unilaterally extending the grace period covering checks on agri-foods moving from Great Britain to Northern

142 Martin Fletcher, 'The Northern Ireland riots have exposed Boris Johnson's reckless complacency' (*The New Statesman*, 12 April 2021) <https://www.newstatesman.com/politics/northern-ireland/2021/04/northern-ireland-riots-have-exposed-boris-johnson-s-reckless> accessed 16 April 2021.

143 Clare Rice, 'Free us': the DUP's Northern Ireland Protocol strategy' (*London School of Economics*, 4 February 2021) <https://blogs.lse.ac.uk/brexit/2021/02/04/free-us-the-dups-northern-ireland-protocol-strategy/> accessed 16 February 2021.

144 ibid.

Ireland.[145] The violence began as a result of an oversight by both the UK and the EU, however. The two parties committed to preserve peace but did so in a manner which, in the opinion of loyalists, went against the principles of parity of esteem[146] and consent[147] enshrined in the Good Friday Agreement. Indeed, some members of the Democratic Unionist Party (DUP), Northern Ireland's largest unionist party, have prepared legal challenges against the Protocol for this very reason.[148]

All of this relates to Article 18, the provision of the Protocol which is supposed to protect the spirit of the Good Friday Agreement through "democratic consent". However, the vote which the Northern Ireland Assembly is supposed to have on the continuing operation of the Protocol is not scheduled until 2024. Considering the volatile atmosphere in Northern Ireland as a result of the Protocol,[149] it seems that the best solution would have been to organise a referendum to allow the electorate to have their say on the issue, although the more likely solution now is for the EU-UK joint committee to work on finding a solution together.[150] It seems counter to the spirit of democracy to change the way the economy and politics of a region work *before* giving its citizens an opportunity to vote on it. That is the fault of the UK government[151] and not the EU, but they did jointly agree on the Protocol. Indeed, the piece of secondary legislation[152] regarding the vote on the Protocol was the subject of a legal challenge brought in the High Court in Belfast, but which is expected to end up in the UK Supreme Court.[153] The applicants in the legal challenge argue that the secondary legislation

145 Alberto Nardelli, 'EU to Delay Brexit Legal Action Amid Northern Ireland Violence' (*Bloomberg*, 9 April 2021) <https://www.bloomberg.com/news/articles/2021-04-09/eu-to-delay-brexit-legal-action-amid-northern-ireland-violence> accessed 18 April 2021.

146 This means that the unionist-loyalist community and the nationalist-republican community must be treated equally and that a majority of voters in both communities must approve of a measure which would have constitutional consequences.

147 This means that fundamental constitutional change in Northern Ireland can only occur through a referendum. The Good Friday Agreement refers to the consent principle in the context of Northern Ireland leaving the UK and uniting with the Republic of Ireland, though Unionists argue this should also extend to the Protocol.

148 'DUP leadership starts challenge against Northern Ireland protocol' *The Guardian* (London, 21 February 2021) <https://www.theguardian.com/uk-news/2021/feb/21/dup-leadership-starts-legal-challenge-against-northern-ireland-protocol> accessed 18 April 2021.

149 Jonathan Powell, 'Peace in Northern Ireland is in danger- Johnson's lies and inaction offer no help' *The Guardian* (London, 11 April 2021) <https://www.theguardian.com/commentisfree/2021/apr/11/boris-johnson-posturing-has-put-northern-ireland-fragile-peace-at-grave-risk> accessed 18 April 2021.

150 Nardelli (n 145).

151 It is the responsibility of the UK government to decide how Northern Ireland gives consent. The process is legislated for by the Protocol on Ireland-Northern Ireland (Democratic Consent Process) (EU Exit) Regulations 2020.

152 ibid.

153 McBride (n 77).

alters "constitutional statutes"[154] such as the Acts of Union 1800[155] and the Northern Ireland Act 1998[156] by removing the cross-community voting mechanism central to the Good Friday Agreement. The plaintiffs argue on five grounds that the Protocol is unlawful: first, that the terms of the Protocol violate the Acts of Union 1800; second, that the Protocol conflicts with the Northern Ireland Act 1998; third, that Article 18 of the Protocol is incompatible with the usual provisions for cross-community voting in the Northern Ireland Assembly (Article 18 mandates a simple majority vote); fourth, that the Protocol is incompatible with Article 3 of the European Convention on Human Rights (ECHR), as Northern Ireland has no way of having any voice in the creation of EU law, yet has to follow it; and fifth, that the Protocol breached Article 50 TEU by providing for the continued application of EU law outside the EU.

The challenge brought by Traditional Unionist Voice (TUV) leader Jim Allister in the High Court of Northern Ireland was rejected, as was the appeal to the Northern Ireland Court of Appeals.

The case will proceed to the United Kingdom Supreme Court. Keegan LCJ of the Northern Ireland Court of Appeals identified the following as the legal questions for the UK Supreme Court to consider: whether the Court of Appeal erred in law by concluding that (a) Article 6 of the Acts of Union did not prevent the UK Government from effecting the Withdrawal Agreement and (b) that the European Union Withdrawal Act 2018 lawfully modifies Article 6; whether the Court of Appeal erred in law by failing to conclude that the modification of Article 6 constitutes a change in the constitutional status of Northern Ireland, in conflict with the Northern Ireland Act 1998; and whether the Court of Appeal erred in law by concluding that the Protocol lawfully disapplied section 42 of the Northern Ireland Act 1998.[157]

Allister[158] has the potential to be a landmark case in UK constitutional law; it may mark the first time any court in the UK has had to resolve a conflict between two constitutional statutes. A ruling on the case is not expected until 2023.

There is the argument that the use of the joint committee to resolve these issues is undemocratic in itself: representatives of Northern Ireland did not have a voice in the Protocol which has had profound economic, social, and political

[154] *Thoburn v Sunderland City Council* [2002] EWHC 195 (Admin) [62] (Laws LJ). In this case, Laws LJ stated that 'constitutional statutes' such as the Acts of Union are immune from implied repeal, which is a concept in UK constitutional law which holds that where two pieces of legislation contradict each other, the latter Act takes precedence. This principle was approved by the UK Supreme Court in *BH v The Lord Advocate (Scotland)* [2012] UKSC 24 [30] (Lord Hope).

[155] 40 Geo. 3 c.38.

[156] c. 47.

[157] Alan Erwin, 'Unionist challenge to NI protocol to proceed to UK's Supreme Court' *The Irish Times* (Dublin, 25 April 2022) <https://www.irishtimes.com/news/crime-and-law/courts/criminal-court/unionist-challenge-to-ni-protocol-to-proceed-to-uk-s-supreme-court-1.4861355> accessed 24 August 2022.

[158] *Re Jim Allister's application for Judicial Review* [2021] NIQB 64.

effects. Before adopting the Lisbon Treaty, most member states offered their citizens a chance to vote on it. The voters of Ireland rejected it, then gained concessions which then led to the Irish electorate approving the Treaty.[159] The Treaty establishing a Constitution for Europe was not adopted because French and Dutch voters rejected it in referenda.[160]

The point of all of this is that it is usually the European way to offer citizens a chance to vote on issues such as these. Article 2 TEU provides that the EU is founded on, amongst others, the value of respect for democracy. It seems counter to this principle the EU was founded on to change the way the economy of Northern Ireland works without giving the people of Northern Ireland a say.

The violence in Northern Ireland has not been caused by the Protocol alone,[161] but the decision to hold a referendum on it before it entered into force would have at least allowed it to enjoy democratic legitimacy. It is good that even under strained relations the joint committee is able to produce solutions to the problems the Protocol has been going through;[162] however, the need for the people of Northern Ireland to have their say sooner rather than later has become very clear.

V. Conclusion

This article has attempted to put forward the case that the Northern Ireland Protocol is in some regards untenable. It has created frictions in trade between Northern Ireland and Great Britain, to the detriment of all members of the community in Northern Ireland. It has pushed Northern Ireland to establish closer trade links with Ireland and the EU. The Protocol has been the cause of much social unrest. Furthermore, its policy on goods moving between Britain and Northern Ireland may protect the single market, but it is disproportionate given that less than 1 per cent of trade between the EU and third countries passes between Britain and Northern Ireland. Whether the CJEU will have much of a role in Northern Ireland is also hard to tell—the current British government has

159 'Ireland backs EU's Lisbon Treaty' (*BBC News*, 3 October 2009) <http://news.bbc.co.uk/1/hi/8288181.stm> accessed 18 April 2021.

160 Patrick Wintour, 'EU scraps timetable for ratifying constitution' *The Guardian* (London, 17 June 2005) <https://www.theguardian.com/politics/2005/jun/17/eu.politics> accessed 18 April 2021.

161 According to police sources, the violence was stirred up by criminal gangs in response to a crackdown on their activities. The decision not to prosecute those who organised and attended the funeral of a leading member of the IRA in breach of Covid regulations also played a role. But the belief that the UK government has abandoned Northern Ireland is what is causing a lot of the discontent in the loyalist community. See 'The Guardian view on the riots in Northern Ireland: situation dangerous' *The Guardian* (London, 6 April 2021) <https://www.theguardian.com/commentisfree/2021/apr/06/the-guardian-view-on-the-riots-in-northern-ireland-situation-dangerous> accessed 18 April 2021.

162 Editorial comments, 'Sour lessons from the Union's first encounters with the UK as a 'free and sovereign country' (2021) 58(1) Common Market Law Review 1, 9.

shown a propensity for disregarding its obligations under international law, whether as a negotiating tactic or not. At the end of April 2021, the European Commission paused the legal action it was intending to bring against the UK with an interest in finding a solution to Northern Ireland's current problems. The latest talks will involve Northern Ireland businesses in an attempt to reach a breakthrough on trading arrangements.[163] While this may not solve the current unrest in Northern Ireland completely, given that it was caused by more than just the Protocol, it makes for a refreshing change in how EU-UK policy on Northern Ireland has been decided. History has shown that the Northern Irish, regardless of political affiliation, do not appreciate the future of their region being decided without them having a role to play. The recent violence and tensions should be a wakeup call to the EU and the UK about the reality of the situation.

[163] Naomi O'Leary, 'EU seeks business input over problems with Northern Ireland protocol' *The Irish Times* (Dublin, 17 April 2021) <https://www.irishtimes.com/news/world/europe/eu-seeks-business-input-over-problems-with-northern-ireland-protocol-1.4539750> accessed 19 April 2021.

Cambridge Law Review (2022) Vol VII, Issue 2, 28–55

'Legitimate' Protest in European Human Rights Law: A Critical Reconstruction

MYTHILI MISHRA*

ABSTRACT

This article studies the construction of 'legitimate' protest in European human rights law. It uses the jurisprudence of the European Court of Human Rights to understand and evaluate what kinds of protest the Court legitimises, and what kinds it does not. The conceptual map consists of three ideas: responsibility, disruption, and offence. It is argued that these three fundamental strands come together to construct the Court's account of 'legitimate' protest. This account is also reconstructed through a critical evaluation of the Court's justifications, enabling us to interrogate the Court's judgments and criticise them for inadequately protecting the right to protest. It concludes with observations about what the findings mean for the protection of human rights and democracy, positing that the Court offers only limited or no protection to protestors who do not fit a certain model, which is a threat to democracy.

Keywords: Right to protest, ECHR, freedom of expression, freedom of assembly, democracy

I. INTRODUCTION

We have been living through an age of pandemic, but also of protest. From the global Black Lives Matter (BLM) protests to women's rights demonstrations in Poland challenging the tightening of anti-abortion laws, people took to the streets. All of this occurred in the middle of an unprecedented global pandemic which posed new threats to the health and safety of protestors and gave new

* Advocate (India), LLB (London School of Economics and Political Science), BA (Lady Shri Ram College, Delhi University). The author is grateful to Dr Sarah Trotter and Professor Kai Möller for their feedback and guidance.

opportunities to governments to crack down on dissent. Although there is nothing new about protests, these raised some novel questions and sparked renewed debate on old ones. Should BLM protestors be allowed to topple racist statues? Should women's rights activists be allowed to protest during a public health emergency?

This article is an attempt to understand how European human rights law answers these questions, and to interrogate those answers. In particular, it seeks to study how the jurisprudence constructs 'legitimate' protest and punishes 'illegitimate' protest. Why European human rights law? First, the European Convention on Human Rights (ECHR) legal order has a rich body of jurisprudence going back decades on this topic. Second, the European Court of Human Rights (ECtHR) is an international human rights court, so unlike domestic law it can show the 'bigger picture', and since its *raison d'être* is protection of human rights, it is expected to provide the highest possible protection for rights.

In theory, the right to protest is guaranteed under Articles 10[1] and 11.[2] Yet, as a discussion of the case law will show, "[i]t is not so much a right to protest, rather a *fight* to protest, because … the law is weighted against the protestor."[3] Preliminarily, it must be noted that both these rights are qualified rights and can be limited in light of public safety, protection of health or morals, rights and freedoms of others, etc. Article 11 additionally only protects peaceful assembly, and thus any protest not adjudged to be 'peaceful' is outside its scope. Evidently, these concepts are not self-explanatory, and much depends on how they are interpreted.

[1] Freedom of expression:

1. Everyone has the right to freedom of expression. This right shall include freedom to hold opinions and to receive and impart information and ideas without interference by public authority and regardless of frontiers. This Article shall not prevent States from requiring the licensing of broadcasting, television or cinema enterprises.
2. The exercise of these freedoms, since it carries with it duties and responsibilities, may be subject to such formalities, conditions, restrictions or penalties as are prescribed by law and are necessary in a democratic society, in the interests of national security, territorial integrity or public safety, for the prevention of disorder or crime, for the protection of health or morals, for the protection of the reputation or rights of others, for preventing the disclosure of information received in confidence, or for maintaining the authority and impartiality of the judiciary.

[2] Freedom of assembly and association:

1. Everyone has the right to freedom of peaceful assembly and to freedom of association with others, including the right to form and to join trade unions for the protection of his interests.
2. No restrictions shall be placed on the exercise of these rights other than such as are prescribed by law and are necessary in a democratic society in the interests of national security or public safety, for the prevention of disorder or crime, for the protection of health or morals or for the protection of the rights and freedoms of others. This Article shall not prevent the imposition of lawful restrictions on the exercise of these rights by members of the armed forces, of the police or of the administration of the State.

[3] John Cooper QC, 'The Right to Protest?' (Speech at 25 Bedford Row) <https://www.25bedfordrow.com/cms/document/The_Fight_to_Protest___John_Cooper_QC.pdf> accessed 23 November 2020 (emphasis added).

Further, the level of protection the ECtHR offers to protestors will seem inconsistent with its position on political expression more generally. The Court has justifiably "afforded a very high level of protection to political speech"[4] and scrutinises restrictions on such speech carefully.[5] It has said in *Wingrove v United Kingdom*[6] that "there is little scope … for restrictions on political speech or on debate of questions of public interest". One would thus expect it to protect protest, which is also a form of political expression, as enthusiastically. However, as this article shows, this is not the case. Moreover, while the jurisprudence on free speech shows that politicians are free to use exaggerated and provocative language to make their point,[7] the case law on protest does not extend the same privilege to the people. This deprives citizens of a voice beyond the ballot, and only protects politicians, who are more likely to be elites. These points will be analysed further below.

This article will show that the ECtHR's idea of 'legitimate' protest consists of three strands, which will be reconstructed through a critical evaluation of the Court's justifications. The three concepts that form the substantive framework are responsibility, disruption, and offence. All three are interrelated and overlapping in some respects. Together, they explain and give an account of the Court's vision. The logic of all three concepts is distinct and may be contradictory. For example, in some cases on responsibility and disruption, the Court carefully emphasises that it is condemning the *act* (such as blocking roads) and not the speech. In contrast, in cases where the Court has protected 'offensive' protests, it has stated that the method and the message are inherently linked. Yet, all three are united by one underlying idea condoned by the Court—protests should not cause too much inconvenience. As John Cooper KC has said, "[t]here is, of course, still a right to protest … just as long as it is … not so noisy and inconvenient as to get in the way of the powerful."[8]

The purpose here is not simply to tell a story, but to question and critique how the Court defines these concepts and applies them. For instance, there is a responsibility to not engage in violent protest, but how should violence be construed? Would a broad definition of violence limit the kinds of permissible protest? What does this mean for human rights and democracy? The following sections ask and attempt to answer such questions.

4 Erica Howard, 'Gratuitously Offensive Speech and the Political Debate' (2016) 6 EHRLR 636.
5 ibid 636.
6 *Wingrove v United Kingdom* App no 17419/90 (ECHR, 25 November 1996).
7 Howard (n 4) 637.
8 Cooper (n 3).

II. Responsibility

The first theme in the characterisation of 'legitimate' protest is that of responsibility. While the duty is usually on the guarantor (the State) to ensure that rights are upheld, the Court has emphasised the duties of rights-bearers as well. The text of Article 10 itself says that its exercise "carries with it duties and responsibilities", so this finds textual support. Thus, the responsibility to act or refrain from acting in a certain way can be conceptualised as a duty imposed on the actor. This is fundamentally different from how the protestors may view their actions themselves—they may feel the responsibility *to* protest, which could then justify certain means. However, the protestors' perspective is not the one the law adopts, and this tension will be seen below. It imputes responsibilities *on* them: towards other citizens, fellow protestors, towards oneself, and even property. It will be argued that the extent to which the Applicant acts 'responsibly', according to the Court's understanding, influences whether a violation of their Convention rights is found. Further, how the Court constructs an account of a 'responsible' protestor will be critically examined.

A. Time, place, manner

One aspect of responsibility is time, place, and manner restrictions on protests. The Court has reiterated that some regulation is permissible for the maintenance of public order, but this cannot be used to stifle dissent.[9] The most important judgment on this is *Navalnyy v Russia*,[10] one of the very few Grand Chamber cases on this topic. Mr Navalnyy is a political activist and opposition leader. He brought an application to Strasbourg relating to consistent attempts by the government to restrict his political activity.

In *Navalnyy*, the Court laid down principles to be followed in assessing the right to protest under Article 11 which merit close scrutiny. It first noted that notification/authorisation procedures for protests cannot be an end in themselves. However, they are justified in general, if they do not "encroach upon the essence of the right"[11] and the purpose is to guarantee the smooth conduct of assemblies. This does not seem problematic at first, especially since there is an exception for spontaneous demonstrations. However, the Court did not elaborate on the

[9] This is becoming crucial in today's political climate, when governments are imposing extensive restrictions and using sanctions when these exact procedures are not followed.

[10] *Navalnyy v Russia* App nos 29580/12, 36847/12, 11252/13, 12317/13, 43746/14 (ECHR, 15 November 2018).

[11] ibid at [100].

"essence" of Article 11, and a vague concept such as this enables wide discretion and could lead to serious interferences with the freedom of assembly, legitimised by the Court itself. Furthermore, the ECtHR carved this exception narrowly, to only include spontaneous demonstrations in *special circumstances*, hollowing it out since the Court can decide what constitutes a legitimate reason to demonstrate spontaneously. Here, on the occasions where the Applicant was protesting, the Court emphasised that the gatherings were peaceful and caused hardly any disturbance, so they should have been tolerated. Yet, it immediately restricted this:

> The intentional failure […] to abide by these rules and the structuring of a demonstration, or of part of it, in such a way as to cause disruption to ordinary life and other activities to a degree exceeding that which is inevitable in the circumstances constitutes conduct which cannot enjoy the same privileged protection … as political speech or debate on questions of public interest or the peaceful manifestation of opinions on such matters. On the contrary … Contracting States enjoy a wide margin […].[12]

This highlights the dual responsibility the Court attributes to protestors: they must not intentionally break the notification rules, and they should not cause disruption that is more than what is inevitable. The fact that *this* kind of assembly does not attract the same protection that other kinds do, and the fact that States have a wide margin of appreciation when it comes to this, shows that the Court is carving out a category of illegitimate protest: *intentionally unlawful* and *unnecessarily disruptive*. Disruption will be discussed below; however, for now it must be noted that some kinds of protest are inherently disruptive, and that is their point. The Court saying that this does not attract the same protection as political speech limits freedom of assembly to only some 'acceptable' assemblies. This is troubling since protest is also a form of political expression, warranting a narrow margin. Moreover, saying that it does not deserve the same privileged protection as *peaceful* expression is even more concerning—it seems as though the Court is saying that non-violent but disruptive forms of protest are not peaceful.[13] Since Article 11 only protects peaceful assembly, this excludes them from protection.

Furthermore, the issue with the position on spontaneous assemblies becomes clearer from *Éva Molnár v Hungary*.[14] Here, demonstrators started to

12 ibid at [156].

13 See Helen Fenwick and Gavin Phillipson, 'Direct Action, Convention Values, and the Human Rights Act' (2001) 31 Legal Studies 535.

14 *Éva Molnár v Hungary* App no 10346/05 (ECHR, 7 October 2008).

protest against the statutory destruction of ballots. The Court reiterated that the absence of prior notification can never be a legitimate basis for crowd dispersal. However, it then evaluated whether the special circumstances under which the notification requirement can be set aside were present. First, it noted that the election result was released two months earlier and had been objectively established. Second, if the demonstrators wished to express solidarity with other protestors, the Court was not persuaded that this could not happen with prior notification.

The first objection is logical since it casts doubt on the spontaneity. However, interestingly, the Court added that the election results had been objectively established. Should the factual basis of the topic matter? For instance, can an assembly be 'illegitimate' in the eyes of the Court if the facts are contested, and the assembly is held by vaccine or climate change sceptics? Although there is overwhelming scientific consensus on these issues, making them objectively established, to say that these gatherings are not as protected as others comes close to content-based restrictions.[15] Perhaps there is a responsibility to protest only when the message has some factual basis. Moreover, it is concerning that the Court stated that expressing solidarity with another protest is not spontaneous. Fulfilling the notification requirements may take a few days, even weeks, and the momentum may die by then. This shows why the special circumstances element is problematic—it allows the Court to adjudicate which spontaneous assembly is legitimate and which is not, and this decides the responsibility of the organisers to fulfil the notification requirement. Thus, not only is there a responsibility to follow procedures, but this also varies based on the nature of the protest.

Non-spontaneous assemblies are similarly restricted. In an earlier Chamber decision, *Lashmankin and Others v Russia*,[16] the Court defended the right to choose the time, place, and manner of protesting, especially when these are important to the participants. However, this should be within the limits of restrictions under Article 11 paragraph 2 and there is a wide margin. Thus, the *right* to choose the time, place, and manner also entails the *responsibility* to put up with certain restrictions and deference towards national authorities. The line between a legitimate and an illegitimate assembly can thus be drawn by the government, with the Court adopting a light-touch review. Mead has therefore argued for a narrow margin to protect political speech and democracy when peaceful communicative

15 See Helen Fenwick, 'The Right to Protest, the Human Rights Act, and the Margin of Appreciation' (1999) 62 MLR 491.

16 *Lashmankin and Others v Russia* App nos 57818/09 and 14 others (ECHR, 7 February 2017).

action is concerned.[17] This is consistent with the free speech jurisprudence of the ECtHR.

B. Reprehensible conduct

Following on the 'legitimate' manner of protests, the category of "reprehensible conduct" offers interesting insight. This category has been constructed by the Court, but since there is no definition in the case law, it is a malleable concept. It will be argued here that the understanding of "reprehensible conduct" should be limited to violence, defined as bodily harm to individuals, to ensure that freedom of assembly is adequately protected. An early discussion of reprehensible conduct can be seen in *Ezelin v France*,[18] a Chamber judgment from 1991. The Applicant, a lawyer, took part in a protest where the police claimed the protestors shouted slogans, painted 'offensive' graffiti, and turned violent. Disciplinary sanctions were imposed on him by the Bar. The Court found a violation of Article 11; there was no evidence that *he himself* made threats or daubed graffiti, and

> [F]reedom to take part in a peaceful assembly […] is of such importance that it cannot be restricted in any way […] *so long as the person concerned does not himself commit any reprehensible act* […].[19]

This shows an element of individual responsibility in the Court's reasoning, focusing on what the protestor did. It also means that there is a category of acts that are reprehensible, showing normative condemnation of such behaviour, and making the method of protest illegitimate. Although the Court does not explicitly clarify whether it considers the painting of graffiti to be reprehensible, disassociating the Applicant from the act shows some disapproval, and conversely, approval of the Applicant as a 'responsible' protestor. It may be asked whether painting political slogans on public property is worthy of being deemed reprehensible, since it does not hurt individuals and can be a form of political expression. However, for now, the relationship between individual responsibility, reprehensible conduct, and the illegitimacy of protest is important for further consideration.

17 David Mead, *The New Law of Peaceful Protest: Rights and Regulation in the Human Rights Act Era* (Hart Publishing 2010).

18 *Ezelin v France* App no 11800/85 (ECHR, 26 April 1991).

19 ibid at [53] (emphasis added).

The leading case on this is *Kudrevičius and Others v Lithuania*,[20] a controversial Grand Chamber case in which Lithuanian farmers protested against the fall in wholesale prices and a lack of subsidies by blocking three major roads. The District Court found them guilty of incitement to or participation in rioting, concluding that their actions had to be characterised as a riot, and they were given custodial sentences. The Court unanimously found no violation of Article 11. First, the judgment clarified that this was not a violent assembly, since vehicles had been used to block highways, not to cause bodily harm. Yet, secondly, it noted that disruption of traffic was intentional to attract attention and that "purposely obstructing traffic and the ordinary course of life in order to seriously disrupt the activities carried out by others is not at the core of … Article 11."[21] Third, it focused on the means used which had "shown a severe lack of respect"[22] for others by restricting public life to a greater extent than freedom of assembly should normally do. Finally, these actions were taken in a context where they were already given permission to demonstrate elsewhere, there were ongoing negotiations, and courts remained an alternative. Therefore, these acts constituted "reprehensible acts" which may justify criminal penalties.

This case is significant for what the Court said, but also what it did not say. It did not discuss the fact that domestic courts convicted the farmers of rioting, which is *by definition* violent, a claim dismissed by the Court itself. Furthermore, it stated comparative European Union (EU) law in *Schmidberger*,[23] where a motorway was blocked by protestors and the European Court of Justice (ECJ) found in favour of the right to protest, without discussing it. What it did say is even more problematic. In characterising disruption of traffic as "reprehensible conduct", the Court went beyond what the Convention says—Article 11 only protects non-violent assembly, but there is no value judgment on the use of other means. Commentators have similarly argued that the Court broadened the definition of "reprehensible acts" beyond acts such as violence and included the much less severe act of creating roadblocks, dangerously widening the scope for intervention, which leads to the "criminalisation of social protest".[24]

Further, to say that these means show disregard for the lives of others equates a traffic-free commute with a human right. Here, the right to protest is competing against lesser interests, themselves not worthy of protection as rights,

[20] *Kudrevičius and Others v Lithuania* App no 37553/05 (ECHR, 15 October 2015).

[21] ibid at [97]. See further discussion on this in the next section.

[22] ibid at [131].

[23] Case C-112/00 *Eugen Schmidberger, Internationale Transporte und Planzüge v Austria* ECLI:EU:C:2003:333.

[24] Ella Rutter and Jasmine Rayée, '"Do You Hear the People Sing?": *Kudrevičius v Lithuania* and the Problematic Expansion of Principles that Mute Assemblies' (*Strasbourg Observers*, 2 December 2015) <https://strasbourgobservers.com/2015/12/02/do-you-hear-the-people-sing-kudrevicius-v-lithuania-and-the-problematic-expansion-of-principles-that-mute-assemblies/> accessed 16 January 2021.

which defeat it.[25] Traffic, an (often minor) inconvenience, is equated with, and trumps, a human right. It can be argued that sometimes the competing interests are rights themselves, such as the right to health being engaged if ambulances are unable to pass. However, the argument here is for *a* right to disruptive protest, not an absolute one. The facts of this case do not show instances of ambulances being blocked. The Court primarily relied on the delays and disruption experienced by trucks and other vehicles, which is inadequate for a blanket prohibition of roadblocks.

Moreover, the Court's analysis of alternatives was ahistorical and apolitical. The protest was organised in the first place since negotiations were not working, one reason for which could be the imbalance of power between the State and farmers,[26] who do not have the resources and influence that corporations do. Fenwick has also argued that such methods may provide the only avenue available to marginalised groups if they wish to participate in democracy, since the democratic process is exclusionary.[27] Suggesting that bringing legal action was a viable alternative is also questionable, since litigation is an option in nearly every case, so there would be no need for protests. The protest route can arguably be more accessible and effective than courts.

Nevertheless, *Kudrevičius* remains a key case. What it tells us about responsibility and the legitimacy of protest is significant. The emphasis is less on individual responsibility and more on the collective responsibility of demonstrators. They have responsibilities towards fellow citizens and must not restrict public life to a greater extent than one should 'normally' do. Here, the Court is delineating what a legitimate amount of restriction of ordinary life is, which makes it normal.[28] This is not based on frequency, since roadblocks are not extraordinary, but on a *normative* understanding of where the line needs to be drawn. Further, paradoxically, protestors must engage with other means available to them (negotiations or litigation) and not intentionally disrupt the lives of others to attract attention. The Court was not sympathetic to the argument that the blockade was needed as a last resort, or even as a political act coexisting with negotiations. Thus, these are important lessons in which responsibilities must be followed for a protest to be 'legitimate' at Strasbourg.

Violence is supposed to be the clearest case of reprehensible conduct, but the way in which it has been defined is questionable. *Razvozzhayev v Russia and*

[25] Mead (n 17) 95.

[26] The 2020-2021 farmers' protest in India, which lasted over a year, and culminated in the repeal of the farm laws which were being opposed, shows the radical potential of protests in giving a voice to those at the periphery of electoral politics and in achieving political goals.

[27] Fenwick (n 15) 493.

[28] It must be noted that a vision of 'ordinary life' is also being constructed here.

Ukraine[29] offers a good starting point. In this case, the Applicants were convicted of organising mass disorder during a demonstration. The Court held that since the first Applicant intentionally led some individuals to break through the police cordon, and this triggered clashes, his conduct fell outside the notion of peaceful assembly. His application was therefore dismissed as incompatible with the Convention. This raises several questions. First, can breaking the police cordon be fairly labelled as a violent act? It disobeys police orders and may even create disorder, but it does not necessarily cause bodily harm. Additionally, the Applicant did not have violent intentions.[30] However, this is not how the Court saw the notion of violence here, diverging from *Kudrevičius*.

Second, it appears as though the Court is suggesting that the first Applicant's actions were not peaceful since they led to *others* following suit, which triggered clashes. However, this contradicts what the Court said in *Ezelin* about individual responsibility, which implied that assemblies are not violent, people are.[31] A case could be made for incitement to violence, or leading others to engage in violence, but this was not considered. The minority opinion[32] further highlighted that the clashes had occurred due to the accumulation of many people and crowd control measures by the police. Crucially, they opined that the mere fact that someone "contributed to the onset of clashes" does not necessarily exclude that the person acted peacefully—context is important. However, on the majority's reading, a demonstrator has the responsibility to not act in a way that may motivate others to cause disorder. This is a broad reading of the responsibility of demonstrators (and they must carefully evaluate their actions and their effects), an even broader reading of the meaning of violence, and a narrow reading of the responsibility of the police.[33]

Finally, interesting insight is offered by the relationship between property rights and "reprehensible conduct". *Taranenko v Russia*[34] is a case concerning

29 *Razvozzhayev v Russia and Ukraine and Udaltsov v Russia* App nos 75734/12 and 2 others (ECHR, 19 November 2019).

30 Beril Önder, "Peaceful assembly' and the question of applicability of Article 11' (*Strasbourg Observers*, 17 January 2020) <https://strasbourgobservers.com/2020/01/17/peaceful-assembly-and-the-question-of-applicability-of-article-11/> accessed 17 January 2021.

31 Mead (n 17) 68.

32 *Razvozzhayev* (n 29) Opinion of Judges Lemmens, Yudikivska, and Keller.

33 Contrast this with how the Court justifies the means used by the police during protests, which has gained much attention after the recent BLM protests. In *Primov and Others v Russia* App no 17391/06 (ECHR, 12 June 2014), the Court allowed "the use of special equipment and even firearms," even if some police officers acted contrary to the rules, in response to a roadblock. Similarly, in *Oya Ataman v Turkey* App no 74552/01 (ECHR, 5 December 2006), the Court condoned the use of tear gas, even though it can lead to various medical complications and serious illnesses. In this case it was used merely because of a failure to disperse. This shows how easy it is to justify the use of force, even lethal means, by police officers, while the demonstrators must strictly adhere to their responsibilities related to peaceful assembly.

34 *Taranenko v Russia* App no 19554/05 (ECHR, 15 May 2014).

public property, where violating property rights was seen as reprehensible conduct by the Court. Including a responsibility to respect property rights further expands the category of reprehensible conduct, making more kinds of protest 'illegitimate'. *Appleby v United Kingdom*[35] further gives an idea of how the Court views the relationship between private property and reprehensible conduct, although the ECtHR did not use the term. This case involved a town centre owned by a private company, where the Applicant and others set up stands, displayed posters, and collected signatures. The Court found no violation of Article 10 regarding their removal. It held that freedom of expression had to be balanced with property rights. However, in light of shrinking public space (privatisation of land that was once publicly owned[36]), the nature of the town centre as quasi-public (not entirely private like someone's garden), and the fact that the local authority had used it to promote political proposals (and thus the selective stifling of speech), the State's positive obligations should have been triggered to facilitate the right to protest. Sanderson has further shown how the State is directly responsible—although ordinary management had passed, responsibility for the administration of fundamental freedoms had not, since there were no hands able to take this burden.[37] Yet, as Judge Maruste's Partially Dissenting Opinion said: "the property rights of the owners … were unnecessarily given priority over … freedom of expression and assembly"[38]—even though holding these as equals in abstract terms is erroneous[39] in the first place. Finally, there is scope for potential misuse of rights by powerful private parties,[40] who can limit the forum and topics of discussion.[41] This disproportionately affects the marginalised, who cannot seek alternatives such as the media.[42]

The above cases show the priority attached to property over protest. Dissent must be respectful of the property rights of the State and corporations, and interference with them can dilute the legitimacy of the protest. Thus, there is a responsibility to maintain property rights. One can justifiably ask, "if almost all land is privately owned, where would demonstrators go?"[43] Furthermore, what happens if the protest is against property rights itself? There is no jurisprudence

[35] *Appleby v United Kingdom* App no 44306/98 (ECHR, 6 May 2003).
[36] David Mead, 'Strasbourg Succumbs to the Temptation "to Make a God of the Right to Property": Peaceful Protest on Private Land and the Ramifications of *Appleby v UK*' (2003) 8 Journal of Civil Liberties 98.
[37] MA Sanderson, 'Free Speech in Public Places: The Privatisation of Human Rights in *Appleby v UK*' (2004) 15 King's Law Journal 159.
[38] *Appleby* (n 35) Partly Dissenting Opinion of Judge Maruste.
[39] Mead (n 36) 103.
[40] ibid 107–108.
[41] Jacob Rowbottom, *Democracy Distorted* (Cambridge University Press 2010).
[42] David Mead, 'A Chill Through the Back Door? The Privatised Regulation of Peaceful Protest' (2013) Public Law 100.
[43] Mead (n 17) 74.

on this, but it is possible that, following the discussion above, the Court would say that protestors can take their disagreement elsewhere and express themselves without interfering with the property rights they oppose.

C. Paternalism

The final theme is that of paternalism, or restrictions that are allegedly in the interests of the protestors themselves. It will be contended here that the vision of an autonomous protestor is being supplanted by that of the responsible protestor, who exercises their autonomy in a particular way. In *Cissé v France*,[44] undocumented migrants occupied a church to draw attention to the difficulties they were facing, but they were evacuated. Finding no violation of Article 11, the Court held that although the protest was peaceful, after two months it resulted in the deterioration of the hunger strikers' health and sanitary conditions. The paternalism is clear—protestors have a responsibility towards *themselves*, to protect their own health and well-being. The fact that they are autonomous actors who can weigh the health risks associated with different methods of political participation is not considered. There is also something Orwellian about the fact that the dissidents' own welfare is used to justify curbing their freedom. It further rules out analogous methods of protest, such as situations in which the weather becomes unfavourable, which may make the protestors ill. Moreover, this logic can be (and has been) used to curtail protests during the pandemic, in the interests of the protestors' own health and safety, taking away their agency to assess the risks themselves. Interestingly, in *Cissé*, the Court could have also concluded that the State must ensure that the protestors are either well fed or that their demands are discussed, since it has positive obligations regardless of people's immigration status. Instead, it condoned their detention, deportation, and imprisonment—all in the name of a responsibility to oneself.

Austin v United Kingdom[45] is less straightforward. In this Grand Chamber case, the police used kettling[46] on anti-globalisation demonstrators. No announcement was made to the crowd when the cordon was first put in place. Eventually the weather became cold, no food or water was provided, and there was no access to toilets or shelter either. The High Court judge concluded that apart from a real risk of injury and property damage, there were risks from crushing and trampling. The ECtHR held that there was no deprivation of liberty under Article 5. The measure was imposed to isolate and contain a large crowd in

[44] *Cissé v France* App no 51346/99 (ECHR, 9 April 2002).

[45] *Austin v United Kingdom* App nos 39692/09, 40713/09, and 41008/09 (ECHR, 15 March 2012).

[46] Kettling refers to containment within a police cordon.

volatile and dangerous conditions. There was no alternative measure—this was the least intrusive and most effective.

Here, the Court saw an unruly crowd that needed to be 'managed', instead of a large-scale demonstration in which freedom of expression and assembly had to be protected. This again highlights the dichotomy of a responsible versus autonomous protestor. The demonstrators were prevented from protesting because they would hurt *fellow* demonstrators, a paternalistic understanding of the collective. Further, Oreb has argued that it is difficult to state conclusively whether Austin was safer inside or outside the cordon,[47] questioning whether kettling was for the demonstrators' own protection. Moreover, most demonstrators were not violent, and the violence that did occur could have been in response to the cordon itself, becoming a self-fulfilling prophecy: the police kettle demonstrators, who become violent, which is then used to detain them further. Crucially, the cordon was put up based on 'intelligence' of protestors having violent intentions, since previous anti-capitalist protests had led to a breakdown in public order. Not only does this show that the cordon was not in response to *actual* violence, it also suggests content-based restrictions. Ultimately, *Austin* shows that the responsibility of *some* demonstrators towards others can be used to curtail the liberty of *all*. However, if the Court's "reprehensible conduct" jurisprudence is applied, then only those who engage in violence should be kept in a cordon,[48] instead of effectively bringing the demonstration to an end.

Hence, it can be seen from the above discussion that the Court's construction of the 'responsible' protestor is also rooted in paternalism—the autonomy of the actor is curtailed, allegedly in their own interests, or in the interests of their fellow demonstrators. Certain forms of protest are consequently illegitimate in Strasbourg jurisprudence, not because they harm or affect *others*, but because they apparently harm the protestors themselves.

This section has shown three ways in which the case law has constructed the idea of responsibility. Following these makes a protest 'legitimate' in Strasbourg jurisprudence. We can also see an idea of the 'responsible' protestor emerging from these—one who does not cause too much inconvenience to others. This idea is also prominent in the Court's understanding of disruptive protest, discussed next.

[47] Naomi Oreb, 'Case Comment: The Legality of 'Kettling' after *Austin*' (2013) 76 MLR 735.

[48] David Mead, 'The Right to Protest Contained by Strasbourg: An Analysis of *Austin v. UK* & The Constitutional Pluralist Issues It Throws Up' (*UK Constitutional Law Association*, 16 March 2012) <https://ukconstitutionallaw.org/2012/03/16/david-mead-the-right-to-protest-contained-by-strasbourg-an-analysis-of-austin-v-uk-the-constitutional-pluralist-issues-it-throws-up/> accessed 25 January 2021.

III. DISRUPTION

The second idea in the construction of 'legitimate' protest in Strasbourg jurisprudence is that of disruption. In some sense, the right to protest can be conceptualised as the right to disrupt, since all protests are disruptive in a way—they disrupt the world as we know it.[49] More specifically, forms of direct action *aim* to disrupt—either to attract attention by disrupting the day-to-day functioning of society,[50] or by directly stopping an activity that is seen as unjust (for example, occupying an arms factory to prevent the arms from reaching warzones). This brings the right to protest in tension with other interests, such as economic or security-related interests.

The ECtHR, however, has largely not been sympathetic to disruptive protest. Although it has said that a certain level of disruption is inevitable and the authorities must show a degree of tolerance,[51] the actual level of disruption tolerated in the jurisprudence is minimal. The Court has often seen disruption as a reasonable justification for curtailing the right to protest. Mead's content study of the case law has also shown that where any form of obstructive activity has been engaged in, even if minor, restrictions have been upheld.[52] Although the Court stresses that the condemnation is limited to the means adopted,[53] it is very difficult to separate the subject matter of the protest (that is, political expression) from its form, which becomes evident from the discussion below.

It will be argued here that if a protest is disruptive according to the Court, it is within Strasbourg's constructed category of 'illegitimate' protest. How the Court delineates the boundary between the kind or amount of disruption that should be tolerated and that which should not, as well as what the underlying conception of the 'ideal' protestor is, will also be examined.

A. INTENTIONAL DISRUPTION

One way to understand the construction of disruption as 'illegitimate' protest in the case law is to look at the focus on intention. The Grand Chamber

[49] Shepherd Mpofu, 'Disruption as a Communicative Strategy: The Case of #FeesMustFall and #RhodesMustFall Students' Protests in South Africa' (2017) 9 Journal of African Media Studies 351.

[50] ibid 354.

[51] *Kuznetsov v Russia* App no 10877/04 (ECHR, 23 October 2008).

[52] David Mead, 'The Right to Peaceful Protest under the European Convention on Human Rights – a Content Study of Strasbourg Case Law' (2007) 4 EHRLR 345.

[53] *CS v Federal Republic of Germany* App no 13858/88 (ECHR, 6 March 1989).

judgment in *Kudrevičius*,[54] discussed previously, is noteworthy here. Blocking roads is inherently disruptive, it stops traffic and may cause some upheaval. How the Court understood and assessed disruption offers another lens to view the judgment. First, the Court said that disruption, in this case, was not a side-effect, but *intentional* action to attract attention and to push the government. This means that although some unintentional disruption can be tolerated, using it as a *tool* makes the protest illegitimate. Contrast this with the Chamber judgment in this case, which said that a certain level of disruption is inevitable and should be tolerated. There seem to be two different visions of disruption here. One is based on threshold—disruption could be minimal or serious, with the former being permissible. The other focuses on intention—the legitimacy of disruptive protest depends on whether the protestors intended it to be as such. The Grand Chamber refers to both when it condemns "intentional serious disruption",[55] but it is unclear if it needs to be cumulative or if either is enough. If intention is enough on its own, then this effectively eliminates disruptive protest from the protection of the Convention, unless it is an unintentional side-effect. Crucially, disruption is often a means to an end—to raise awareness about an issue, to motivate people to act, to get the government's attention, or to get media coverage. In all these cases, it is *intentional*, even though the protestors do not want to harm those whose activities are disrupted.

Second, the Court declared that "conduct purposely obstructing traffic and the ordinary course of life in order to seriously disrupt the activities carried out by others is not at the *core* of … Article 11."[56] However, the concept of the "core" or "essence" of a right is undefined and underexplained, thus remaining essentially pragmatic and unprincipled.[57] In the context of EU fundamental rights, it has also been criticised as arbitrary and meaningless for adopting an "I know it when I see it" logic.[58] As such, it remains an uncertain concept. Yet, if the Court's assertion of disruption not being at the core of Article 11 is taken seriously, it implies a deprioritisation[59] of disruptive protest, placing it outside a sacrosanct core.[60] What is essential to a right and what is peripheral is not obvious, and it involves normative choices. For instance, commercial expression is less protected than political expression, but the justification is that freedom of expression exists to

54 *Kudrevičius* (n 20).
55 ibid at [73].
56 ibid at [97] (emphasis added).
57 Sébastien Van Drooghenbroeck and Cecilia Rizcallah, 'The ECHR and the Essence of Fundamental Rights: Searching for Sugar in Hot Milk?' (2019) 20 German Law Journal 904.
58 Mark Dawson, Orla Lynskey, and Elise Muir, 'What is the Added Value of the Concept of the "Essence" of EU Fundamental Rights?' (2019) 20 German Law Journal 763.
59 Van Drooghenbroeck and Rizcallah (n 57) 906.
60 Pierre Thielbörger, 'The "Essence" of International Human Rights' (2019) 20 German Law Journal 924.

protect democracy, not economic interests. No such principled explanation is offered in *Kudrevičius*, even if the merits of the methodological approach are presupposed. Consequently, to say that one method (that of disruption) is not as central to freedom of assembly, is to cast doubt on its legitimacy as a method of protest—as abnormal or unnecessary, even reprehensible.

Finally, in justifying the custodial sentence, the Court emphasised that such "inconvenience does not seem disproportionate when compared to the serious disruption of public order provoked."[61] It reached this conclusion following a wide margin, due to a focus on disruption, as opposed to the political aspect of the protest, which should have led to a narrow margin. Justifying a custodial sentence in response to "serious disruption" to traffic is concerning, and it shows the extent to which disruptive protest is seen as illegitimate: reprehensible enough to justify *imprisonment*.

B. Disorder

Another trend has been the equation of disruption with disorder. *Steel and Others v United Kingdom*[62] is a paradigmatic case. Here, the Applicants undertook various disruptive forms of protest. The first Applicant took part in a protest against a grouse shoot by attempting to obstruct and distract those taking part. The second Applicant took part in a protest against the building of an extension to a motorway by breaking into the construction site and climbing on trees and machinery.[63] Here, the Court recognised that physically impeding activities constitutes expression under Article 10. However, it noted that regard must be had to the "risk of disorder arising from the persistent obstruction … as [the shooters] attempted to carry out their lawful pastime."[64] There are two noteworthy things here. First, the Court characterised direct action as somewhat inherently disorderly. It may be worth asking how obstructing a shoot can lead to disorder. This was a non-violent protest in which the demonstrators sought to prevent an activity they found morally objectionable. What the risk of disorder was, and how it would manifest itself, was not discussed. It has been contended that disorder has often been used to mean mere inconvenience or annoyance.[65] This seems convincing, as the logical link between obstructing a shoot and an outbreak of

61 *Kudrevičius* (n 20) at [178].

62 *Steel and Others v United Kingdom* App nos 67/1997/851/1058 (ECHR, 23 September 1998).

63 Although this case is from 1998, it is of immense contemporary relevance. For example, environmental activists opposing the building of the HS2 rail link recently occupied tunnels in London Euston and were removed by the police, which is analogous to the factual matrix in *Steel*.

64 *Steel and Others* (n 62) at [103].

65 Mead (n 17) 90–91.

violence (or other forms of disorder) is tenuous here, unless there were some special circumstances (contrary to the facts). Nevertheless, based on the Court's understanding, since preventing disorder is a legitimate aim for restricting rights, this *form* of protest is seen as illegitimate due to its alleged propensity to cause disorder. Further, the threshold for a protest being characterised as disruptive is quite low, and proving a connection based on disorder (defined as the absence of order, turmoil, civil disturbance, or chaos[66]) is not required.

Second, emphasis was laid on the fact that lawful activities were being disrupted, contrasted with the unlawful activity of obstructing them, and thus breaking the law to prevent a lawful activity is cast as illegitimate. However, this emphasis on legality can be questioned. 'Illegal protest' can be effective, since it makes for sensational news and would be covered by elite-controlled media.[67] It can also make it difficult for a system or activity to function.[68] Moreover, some reframing is needed: this was a peaceful protest on a political issue, and on the other side of the balance was a hobby practiced by some. In these circumstances, stripping both the activities of their context and only evaluating them based on their legality, on the assumption that the hunters "were doing nothing (legally) wrong", means exaggerating the claim they have. This is another instance of the argument that has been made previously, that protest is pitted against lesser claims, which ultimately outweigh human rights. After all, if the Court is "defending the rights of hunters … over the rights of free expression or demonstration it ought, at the very least, admit as such."[69] Fenwick and Phillipson, however, have argued that personal autonomy is threatened when an individual's freedom to choose to take part in morally controversial activities, such as hunting, is curtailed through the imposition of others' views. They have also argued that the hunters' freedom of association is at stake.[70] However, this reasoning overlooks the fact that the protestors' autonomy is *also* at stake. As discussed above, the reason political speech is more valued and highly protected when compared to commercial speech, or here, an individual's recreational activity, is due to its roots in democracy. Thus, the autonomy of protestors should outweigh that of hunters.

66 ibid.

67 Kimberly Brownlee, 'Protest and Punishment: The Dialogue between Civil Disobedients and the Law' in Michael Freeman and Ross Harrison (eds), *Law and Philosophy* (Oxford University Press 2007).

68 ibid 263. It must be noted that if the right to disruptive protest is protected too, it would no longer be illegal. However, the disruptive nature of the protest is still likely to garner media attention.

69 David Mead, 'The Human Rights Act – A Panacea for Peaceful Public Protest?' (1998) 3 Journal of Civil Liberties 206.

70 Fenwick and Phillipson (n 13) 545.

This idea of disruption that the majority favoured was also challenged by the dissenting judges.[71] They questioned if disrupting the shoot to defend animal rights was as dangerous as the majority made it seem. This is especially important since the connection with disorder is not self-evident, and a fundamental freedom is being weighed against a pastime. Nevertheless, this case shows that the ECtHR is unwilling to protect more obstructive and disruptive forms of protest.[72] Thus, very few Strasbourg cases succeed when they are outside the paradigm of peaceful demonstrations and processions.[73] This can be seen as the archetype of 'legitimate' protest, and the further the protestors deviate from it, the more the restrictions on their protest are justified for the Court.

C. Legitimacy of non-disruptiveness

The final set of cases show how the Court has bolstered the legitimacy of certain kinds of protest by characterising them in opposition to disruptive protest, which makes them 'worthy' of protection. Thus, even though the Court found violations in the cases discussed below, it did so by distinguishing them from other, 'illegitimate' forms of protest.

Solo demonstrations, a matter of discussion in *Novikova and Others v Russia*,[74] offer interesting insight here. The Court held that solo demonstrations fall under Article 10. Noting that they are a form of political expression, and that here the demonstrators merely held posters, the Court found that the actions of the police were disproportionate due to multiple factors, and the one relevant for our discussion is "swift termination". The Court stressed that there was only one participant, so prior notification would not have enabled the authorities to take measures to minimise disruption. Further, the ECtHR emphasised that the events were peaceful, and that there was no violence or obstruction of traffic. Thus, an attempt was made to characterise solo protests as harmless and causing no disruption—an attempt to lend legitimacy to them and characterise the demonstrators as 'ideal' protestors. The corollary of this would be that a protest that is disruptive of traffic, for instance, would not enjoy the same protection. If a solo protest involved standing on a zebra crossing with a poster, stopping the flow

[71] According to them, "[w]hat is not … debatable is that to detain for forty-four hours and then sentence to twenty-eight days' imprisonment a person who, albeit in an extreme manner, jumped up and down in front of a member of the shoot to prevent him from killing a feathered friend is so manifestly extreme, particularly in a country known for its fondness for animals ..." (Joint Partly Dissenting Opinion of Judges Valticos and Makarczyk).

[72] Mead (n 52) 356–357.

[73] ibid 359.

[74] *Novikova and Others v Russia* App nos 25501/07, 57569/11, 80153/12, 5790/13, and 35015/13 (ECHR, 26 April 2016).

of traffic to bring the attention of drivers to an issue, it could have been construed differently. Notably, the threshold for disruption of traffic may well be below the kind of scenario in *Kudrevičius*.[75] A single person obstructing traffic may also be characterised as 'disruptive' protest.

Similarly, *Mariya Alekhina and Others v Russia*[76] also illustrates how 'legitimate' protest is cast in opposition to disruption. This case will be discussed in depth in later, but briefly, the facts involved members of a feminist band Pussy Riot performing songs with political messages in a cathedral. One of the factors the Court emphasised in assessing their application under Article 10 was the fact that they did not disrupt any religious services—conduct characterising 'ideal' protestors. Thus, as in *Novikova*, the claim of the protestors was strengthened by disassociating them from an 'illegitimate' form of protest (that is, disrupting services in the cathedral). Had they started their performance in the middle of a service, interrupting the ordinary functioning of the place of worship, the Court could have been less sympathetic.

This section has discussed the construction of disruption as an attribute that makes protest 'illegitimate' at Strasbourg. It has shown how an attempt has often been made to divorce the *means* of disruption from the political *message* in the reasoning of the Court. The ECtHR often reminds the protestors that there were other, non-disruptive, means available. As Mpofu has argued, this can be understood as a plea to protest within the confines of the law, which can be shorthand for protesting on the margins and not disturbing the status quo.[77] This can be seen in the case law as well—so long as the protest does not cause too much inconvenience it is protected, since some disruption is inevitable, but not when it gets too disruptive. Thus, what is protected is only the right to protest in a certain, limited way. This is further seen in the next section on offence, where the means are questioned not only for causing some tangible inconvenience, but mere intangible offence, and the content also comes under scrutiny.

IV. Offence

The final concept in the construction of 'legitimate' protest is offence. The Court has firmly said that if disagreement were enough to prohibit certain kinds of protest, then "society would be … deprived of the opportunity of hearing differing

75 As stated previously, in *Kudrevičius*, vehicles were used to block major roads for a sustained period.

76 *Mariya Alekhina and Others v Russia* App no 38004/12 (ECHR, 17 July 2018).

77 Mpofu (n 49) 360.

views on any question which offends the ... majority."[78] Following this, mere 'dislike' or 'offence' is not enough to muzzle freedom of expression, and a diversity of views is celebrated as intrinsic to democracy. Moreover, there is a positive reason to protect offensive protest: offensiveness can serve as a "pre-political gateway to future civic engagement"[79] and is a "worthy political tool ... to publicise neglected political issues."[80] This is particularly important for minorities, since the majority usually has the power to define offence.[81] Thus, the argument from democracy and minority rights both support a defence of offensive protest.

It will be argued that although a 'right to offend' is protected under the right to protest, and thus the offensiveness of a protest does not render it 'illegitimate', this is applied inconsistently, which raises a deeper question of what an 'offensive' protest is, and how much offence is too much.

A. 'Right to offend'

In a series of cases on freedom of expression, the ECtHR has reiterated that speech or expressive acts can "offend, shock, or disturb"[82] and still be protected. This has led Fenwick to conclude that the content of a protest rarely excludes it from protection.[83] An early example of this is the 1988 case of *Plattform Ärzte für das Leben v Austria*.[84] Here, counterdemonstrators had shouted and thrown eggs and grass at demonstrators. The Court, under Article 11, unequivocally said that although a demonstration may "annoy or give offence,"[85] the demonstrators must be allowed to protest without fear, otherwise they (and others) will be deterred from expressing their views. A violation was not found on the facts, but this principle is significant.

More recently, it was reaffirmed in *Stankov and the United Macedonian Organisation Ilinden v Bulgaria*,[86] where assemblies commemorating certain historical events important to the Macedonian minority in Bulgaria were banned. The demonstrators were accused of separatism and challenging Bulgaria's

[78] *Stankov and the United Macedonian Organisation Ilinden v Bulgaria* App nos 29221/95 and 29225/95 (ECHR, 2 October 2001) [107].

[79] Ian Reilly and Megan Boler, 'The Rally to Restore Sanity, Pre-Politicization and The Future of Politics' (2014) 7 Communication, Culture, & Critique 435.

[80] Anne Graefer, Allaina Kilby, and Inger-Lise Kalviknes Bore, 'Unruly Women and Carnivalesque Countercontrol: Offensive Humour in Mediated Social Protest' (2019) 43 Journal of Communication Inquiry 171.

[81] Marietta Kesting, 'Changing Visual Politics in South Africa: Old and New Modes of Exclusion, Protest, and Offence' in Anne Graefer (ed), *Media and the Politics of Offence* (Palgrave Macmillan 2019).

[82] *Handyside v United Kingdom* App no 5493/72 (ECHR, 4 November 1976).

[83] Fenwick (n 15) 496.

[84] *Plattform Ärzte für das Leben v Austria* App no 10126/82 (ECHR, 21 June 1988).

[85] ibid at [32].

[86] *Stankov* (n 78).

territorial integrity. The Court found a violation of Article 11, and one reason was that no matter how "shocking and unacceptable"[87] certain words or views may be, a total ban endangers democracy. This is a robust defence of unpopular minority opinions. The Court also made the link with democracy explicit: democracy allows disagreement and debate, hence suppressing 'offensive' ideas threatens it. Thus, propagating an 'offensive' message does not make a demonstration 'illegitimate' in the jurisprudence.

Not only are 'offensive' ideas protected, 'offensive' methods are protected as well. *Fáber v Hungary*,[88] for instance, was about a controversial symbol. It involved a counterdemonstrator waving a "provocative" flag, which the government saw as threatening public peace since it could have led to hostile incidents. The Court firmly stated that the "provocative" nature of the flag, that is, the offence caused by it, constituted freedom of expression and was not enough to justify intervention. Further, it noted that mere display of the flag was not capable of disturbing public order or hampering the demonstrators' freedom of assembly. Crucially, it held that "ill-feelings" or "outrage" cannot represent a pressing social need—a move firmly in favour of 'offensive' methods of protest.

This has also been applied to the manner of expressing dissent in *Stern Taulats and Roura Capellera v Spain*,[89] where the Applicants had set fire to a photograph of the royal couple during an anti-monarchist and separatist demonstration. They were charged with the offence of insulting the Crown and sentenced to 15 months' imprisonment. The Strasbourg Court found a violation of Article 10. It noted that the factors referred to by the Constitutional Court were clearly related to the Applicants' political criticism—the effigy of the King represented the monarch as the Head of the State, while the use of fire and turning the photo upside down symbolised rejection. Thus, the "provocative" events were used to attract media attention and went no further than recourse to a certain permissible degree of provocation to transmit a critical message. This is interesting, since the Court suggested that the category of 'offensive' speech in a protest is acceptable only to a *limited extent*, not *categorically*. This will become relevant while discussing other cases, where perhaps the offensiveness went beyond the 'permissible limit'. For now, it is evident that not only is the 'offensive' content of the protest protected, the method of expressing it is also not enough to make a protest 'illegitimate' for the Court.

[87] ibid at [97].

[88] *Fáber v Hungary* App no 40721/08 (ECHR, 24 July 2012).

[89] *Stern Taulats and Roura Capellera v Spain* App nos 51168/15 and 51186/15 (ECHR, 13 March 2018). NB the full judgment was only available in French, so the official summary of the Court (available in English) was relied on.

Thus, preliminarily, it can be said that 'offensive' protest is protected at Strasbourg to facilitate freedom of expression, minority views, and democratic discourse. Following this, it falls under the category of 'legitimate' protest. However, there may be limits to this, since protection is a matter of degree.

B. Too offensive?

Based on the above discussion, it seems that 'offensive' protest is well protected at Strasbourg. However, the story is more complicated. Other cases show that either the Court has been inconsistent in protecting such protest, or the way in which offence is understood in the jurisprudence is limited, severely restricting the scope of the 'right to offend'. It will be argued that the latter offers a better explanation, and although the Court has been strongly in favour of *some* kinds of 'offensive' protest, others are *too offensive* to merit such protection.

The first category of cases is one where no violation of Convention rights was found, and the reasoning of the Court explicitly stated that this was based on the provocative nature of the protest. *Sinkova v Ukraine*[90] illustrates this. Here, the Applicant fried eggs over the Eternal Flame at the Tomb of the Unknown Soldier. This was filmed and posted online with a message about the wastage of natural gas. She was charged and found guilty of desecration of the tomb. Invoking Article 10, the Applicant submitted that she was protesting against the wasteful use of natural gas and tried to draw attention to the fact that the funds used to maintain eternal flames could instead improve the living standards of war veterans. The Court disagreed. First, it said that she was prosecuted "only" for frying eggs, not the "rather sarcastic and provocative" video.[91] Thus, the conviction was due to her conduct and not for her views. However, this strips the performance of all meaning and context,[92] dismissing it as senseless provocation,[93] whereas it was arguably a part of, and central to, the political message. This is particularly true in the age of social media, where such provocative videos tend to 'go viral' and reach a mass audience, thus sparking debate about the issue. Overlooking these nuances, however, the Court singled out the 'offensive' means as a form of 'illegitimate' protest.

90 *Sinkova v Ukraine* App no 39496/11 (ECHR, 27 February 2018).

91 ibid at [107].

92 Ronan Ó Fathaigh and Dirk Voorhoof, 'Article 10 ECHR and Expressive Conduct' (2019) 24 Communications Law 62.

93 Andra Matei, 'Art on Trial: Freedom of Artistic Expression and the European Court of Human Rights' (*SSRN*, 5 February 2020) <https://ssrn.com/abstract=3186599> accessed 5 March 2021.

Second, the Court emphasised the fact that she had many "suitable opportunities"[94] to express her views or participate in "genuine protests"[95] without breaking the law and insulting the memory of soldiers. This again shows contempt for her protest for being offensive, and stresses that there were inoffensive ways to make the point. The use of terms such as "suitable" and "genuine" suggests that the Court saw these qualities as those that characterise 'legitimate' protest, qualities that the Applicant's protest did not share. The minority challenged this—for them, although "extremely provocative", the right to offend, shock, or disturb is a part of Article 10.[96] Thus, the satirical nature of the protest, featuring exaggeration and distortion of reality to provoke and agitate, had to be considered.[97]

Ó Fathaigh and Voorhoof have argued that the aim of protecting the soldiers' memory should not have outweighed freedom of expression—this was a political performance, concerned a matter of public interest, did not involve violence, and had no intention to insult.[98] However, it can also be argued that mere offence should be filtered out at the legitimate aim stage, instead of being balanced. Möller has argued that for a goal to be legitimate, it must be autonomy-related, and ethical dislike (or offence) must not be accorded weight as 'legitimate' in a political community committed to personal freedom.[99] Letsas has also made a similar argument in the context of expression that offends religious feelings. He has contended that there is no right to not be offended, since offence has no independent moral value. To justify state intervention in response to individual offence is in fact the imposition of one view of the good life.[100] Following this, adopting the Court's view means sacrificing freedom of speech to the protection of the feelings of others.[101]

The second category is where a violation has been found, but only due to the imposition of harsh penalties. In these cases, the Court admitted that the outcome may have been different in the absence of strict punishment, and thus these are not cases where 'offensive' protest is being protected. In *Shvydka v Ukraine*,[102] for instance, the Applicant approached a wreath laid by the President and detached part of the ribbon bearing the words "the President of Ukraine V.F. Yanukovych". This was meant to express her disagreement with his policies. She

94 *Sinkova* (n 90) at [110].
95 ibid.
96 *Sinkova* (n 90), Joint Partly Dissenting Opinion of Judges Yudkivska, Motoc, and Paczolay.
97 ibid.
98 Ó Fathaigh and Voorhoof (n 92) 65.
99 Kai Möller, *The Global Model of Constitutional Rights* (Oxford University Press 2012).
100 George Letsas, 'Is there a right not to be offended in one's religious beliefs?' in Lorenzo Zucca and Camil Ungureanu (eds) *Law, State, and Religion in the New Europe: Debates and Dilemmas* (Cambridge University Press 2012).
101 Matei (n 93) 2.
102 *Shvydka v Ukraine* App no 17888/12 (ECHR, 30 October 2014).

was charged with petty hooliganism, convicted, and sentenced to ten days' administrative detention. Although the Court accepted that her acts constituted political expression, it held that the offence of petty hooliganism was not "manifestly inapplicable" to it, since it concerned "offensive behaviour disturbing political order."[103] For the ECtHR, her protest fell under this since she had "resorted to a provocative gesture likely to disturb or insult."[104] A violation was only found due to the imposition of the harshest sanction. Thus, even a minor act such as detaching a ribbon can be construed as 'offensive' and one that disturbs order, which means that the Court saw her protest as 'illegitimate' in some ways. Judge de Gaetano, however, thought that nothing could justify the conclusion that detaching part of a ribbon amounts to offensive behaviour.[105] The argument made here, however, goes further: *even* if it was offensive, it should be protected.

Finally, *Mariya Alekhina*[106] reiterates this. This case was previously discussed in relation to disruption, now the focus will be on offence. The facts are as follows: the Applicants, members of a Russian feminist punk band named Pussy Riot, performed a song which had political messages relating to criticism of President Putin and the support given to him by the Church; as well as supporting LGBT rights, feminism, and the right to protest, on the altar of Moscow's Christ the Saviour Cathedral. They were charged with the aggravated offence of hooliganism motivated by religious hatred. The District Court found them guilty and gave them prison sentences. The reasons included using "obscene language and insulting words", "showing disrespect for society", etc. Thus, their prosecution was based on feelings of religious offence, reiterated by the government's submissions to the ECtHR regarding the duty to not be "gratuitously offensive" towards religion. As Orlova has highlighted, this is because framing the issue as one of religious speech engages a wide margin.[107] As argued earlier, protest is a form of political expression, which should be highly protected in a democratic society. Protest has an element of action that speech does not, but it still falls under the umbrella of freedom of expression. The action is also a part of the expression—it is political. This should justify a narrow margin.

The Court found a violation of Article 10. It first noted that this was artistic and political expression. However, it again engaged the 'rights' of believers, but whether the right to not be offended is a legitimate aim has been challenged above. Following this, it considered the performance as having violated the accepted rules

[103] ibid at [39].
[104] ibid.
[105] ibid, Separate Opinion of Judge de Gaetano.
[106] *Mariya Alekhina and Others* (n 76).
[107] Alexandra V Orlova, 'Pluralism, Democracy, and the Conflict Within: Challenging the State's Narrative by Artistic Forms of Protest' (2019) 27 University of Miami International and Comparative Law Review 1.

of conduct in a place of worship, principally justifying the imposition of certain sanctions. Thus, a violation was only found due to their imprisonment, and the Court legitimised sanctions for engaging in 'offensive' protest. Importance was given to the sanctity of the religious place, without acknowledging why the venue was chosen—the Applicants had submitted that the Patriarch of the Russian Orthodox Church had used the venue for criticising demonstrations against Putin and announcing support for him. Moreover, Kananovich has stressed that the Church has generally legitimised and stabilised the Russian political regime.[108] Instead, by focusing on the religious aspects of the performance, the women were dismissed as "immoral sinners", not deserving protection.[109] The Court seemed to see protest in a place of worship as *inherently* offensive, and thereby illegitimate.[110] One may ask then, what happens if the point of the protest is to criticise what religion or its representatives stand for vis-à-vis women's rights, LGBT rights, etc., and to raise awareness among believers? Protests are not just against governments, but against *all* power structures. *Mariya Alekhina* suggests that these issues need to be debated outside places of worship, since their presence inside these sacred spaces can invite sanctions.

How can we explain these cases? One explanation is that the Court has simply been inconsistent: it has protected 'offensive' protest in some cases and not in others. This would mean that such protest remains within the category of 'legitimate' protest, but the Court has gotten the answer wrong sometimes. However, the sheer number of cases where this has happened[111] suggests that it is a pattern, not an aberration. Another, more nuanced approach, is that to understand this divergence in the jurisprudence we need to ask how the category of 'offensive' protest is constructed in the first place. In *Stern Taulats*, the Court indicated that there is a *permissible* degree of provocation, and perhaps in the other cases the method or message was too offensive—particularly when religious feelings are involved, as in *Mariya Alekhina*. Therefore, a trend similar to the disruption jurisprudence can be seen—offence is protected, but *only to a limited extent*. There are 'legitimate' kinds of offensive protest and 'illegitimate' kinds. Where the boundary is drawn is unclear, since burning a photo is permissible (*Stern Taulats*) but detaching a ribbon (*Shvydka*) is not. Criminal sanctions aside, a lot may turn on the degree of permissible offensiveness, which is determined by

[108] Volha Kananovich, '"Execute Not Pardon": The Pussy Riot Case, Political Speech, and Blasphemy in Russian Law' (2015) 20 Communication Law and Policy 343.

[109] Alexandra V Orlova, 'Russian Politics of Masculinity and the Decay of Feminism: The Role of Dissent in Creating New Local Norms' (2018) 25 William & Mary Journal of Race, Gender, and Social Justice 59.

[110] As Judge Elósegui said in his Partly Dissenting Opinion, "Article 10 does not protect the invasion of churches."

[111] See *Murat Vural v Turkey* App no 9540/07 (ECHR, 21 October 2014) and *Mătăsaru v the Republic of Moldova* App nos 69714/16 and 71685/16 (ECHR, 15 January 2019).

the State that is the subject of the protest itself. Thus, if there is a 'right to offend' at Strasbourg, it is considerably limited. The 'ideal' protestor can theoretically engage in a somewhat 'offensive' protest and still be protected but should be careful to not be 'too offensive'. In practice, given the uncertainty regarding what would be protected, protestors would refrain from offensiveness altogether—a chilling effect on protest and free speech.

This section has shown how 'offensive' protest is both legitimate and illegitimate in Strasbourg jurisprudence, depending on how offensive the Court may deem the protest, and what it defines as offensiveness. It contrasted the 'right to offend' championed in some cases with others where the protest was seen as too provocative to be protected.

V. Conclusion

The idea of 'legitimate' protest in European human rights law thus lies at the intersection of the Court's understanding of responsibility, disruption, and offence. This article has used these concepts to construct a narrative about the jurisprudence, its underlying assumptions, and normative commitments. It has also critically evaluated the Court's account, since the aim was to reform as much as it was to understand.

It has also shown how, underlying the account of 'legitimate' protest in Strasbourg jurisprudence, is a vision of the 'ideal' protestor. They are a responsible protestor who, for instance, follows time, place, and manner restrictions set by the State, even if it significantly limits the protest. Moreover, they do not engage in any kind of disruptive protest, even if peaceful, since it may be construed as causing disorder. Finally, they are not too offensive or provocative in hurting the feelings of the majority or those in power. So long as individuals conform with this vision, the Court defends their rights. The more they depart from this, the more likely it is that the Court would not find a violation of their rights.

We now have some answers to the questions that were posed in the beginning. For example, when we ask if a women's rights protest will be allowed during the pandemic, ECHR law could answer in the negative, since the protestor has responsibilities towards themselves and their fellow protestors. This, however, does not tell us if it *should* be allowed. Following the analysis above, if this imputed responsibility is rooted in paternalism—and paternalism is generally contrary to autonomy, which we value—we might say that it should be allowed, at least under some conditions. Further, if the toppling of racist statues is construed as 'disruptive' (and we have seen how low the threshold for this is), or 'offensive' to the majority or the State, then the Court would not rule in favour of the protestors.

Yet, we could ask if this *really* causes any disruption at all, and who or what is being disrupted. Equally, we could ask if racist symbols are not more offensive to us as a society than tearing down statues, or alternatively, if this is not what the 'right to offend' entails.

What does all of this tell us? A recurring theme in the analysis of the jurisprudence has been the fact that the Court offers only limited or no protection to protestors who do not fit a certain model. There is a substantial degree of deference towards the State, following a wide margin, and the Court does not ask too many questions. This is concerning, since our human rights are being inadequately protected. Moreover, since political expression is being curtailed, democracy is imperilled. Democratic theorists have often noted that what we see today is only an "illusion of democracy",[112] since decisions are made behind closed doors by unaccountable agents.[113] Further, authoritarianism is rising in Europe under the guise of 'illiberal democracy'. In this regard, protests signify people's resistance to the status quo. These are not extra-democratic, they are a crucial (and sometimes the only) way of "challenging established privileges and shaking the existing institutions."[114]

Thus, restricting protests undermines an integral part of democracy—one that acts as a bridge between representative institutions and those they (seek to) represent. The argument becomes more pressing if we ask who can get elected through the mainstream democratic channels and who remains on the outside—a point that has been made throughout this article with regard to marginalised groups. The undocumented migrants in *Cissé*, for instance, do not have the political power that others do, they have to occupy a church to be heard.

Some may agree with the analysis above but argue that it is only certain kinds of protest which are not being protected, so democracy is not under threat. However, the roadblocks and occupations described in this article are "languages of the unheard."[115] D'Arcy has argued that "militant protest" is often good for democracy when it challenges the power of elites, gives a voice to the directly affected, enhances their power, and is sensitive to democratic values.[116] Other reasons why such protests may be the only avenue available have been discussed above. Therefore, an attack on the right to protest is an attack on democracy itself,

112 Robin Celikates, Regina Kreide, and Tilo Wesche (eds), *Transformations of Democracy: Crisis, Protest, and Legitimation* (Rowman & Littlefield Publishers 2015).

113 ibid 1.

114 ibid 2.

115 Stephen D'Arcy, *Languages of the Unheard: Why Militant Protest is Good for Democracy* (Zed Books 2014).

116 i bid 19–21.

and disruptive and offensive protests must be protected if we are to take democracy seriously.

This is not to say that there is no hope for a progressive jurisprudence. There are also trends within the jurisprudence (for instance, some protection of the 'right to offend') that can be leveraged to challenge other principles and judgments. The privileged protection given to political speech can also be used to give analogous protection to political protest. It is not uncommon to see the Court change its outlook. Thus, engagement with trends in the jurisprudence and the discourse on critique and reform must be kept alive.

Cambridge Law Review (2022) Vol VII, Issue 2, 56–71

A Critical Analysis of the Scottish Government's Draft Gender Recognition Reform (Scotland) Bill and its Adherence to the UN Convention on the Elimination of All Forms of Discrimination against Women

ESTHER HODGES*

ABSTRACT

In March 2022, the Scottish government introduced the draft Gender Recognition Reform (Scotland) Bill. The draft Bill aims to streamline the process for those seeking to obtain a Gender Recognition Certificate and so amend the sex on their birth certificate to the gender with which they identify. Its proposed reforms have attracted significant opposition from some. Drawing on qualitative analysis of submissions to the draft Bill's second public consultation, this article argues that opposition is typically based on a reductive, classical sociological conceptualisation of gender, which understands gender as an immutable binary ordained by nature and contends that trans women are not women. By making it easier for trans women to gain legal recognition for the gender with which they identify, those opposing the draft Bill on these grounds therefore argue that its reforms put the rights and freedoms of cis women at risk. This article explores this contention by critically analysing the draft Bill's adherence to the Convention on the Elimination of All Forms of Discrimination against Women (CEDAW). Setting analysis against a framework of two of the CEDAW's most relevant articles and its General Recommendation 28, it argues that the draft Bill is demonstrably in adherence with CEDAW because of its efforts to reduce discrimination against trans women through means which in no way increase the risk of discrimination against cis women. Drawing on postmodernism, this article elucidates a progressive conceptualisation of gender which contends it is not fixed. It argues the draft Bill,

* LLM Human Rights candidate (University of Edinburgh), MSc (SOAS), BA Hons (Oxon).

and indeed CEDAW, could go further in their efforts to reduce discrimination faced by trans women by reducing their evidential reliance on binary conceptualisations of gender. In so doing, they could encourage greater feminist and queer coalitional work, discouraging efforts to pit women's rights against those of trans people to support the emancipation of all women.

Keywords: CEDAW; gender recognition; postmodernism; trans rights; women's rights

I. Introduction

Introducing the draft Gender Recognition Reform (Scotland) Bill ("the draft Bill") to the Scottish Parliament on 3 March 2022, Cabinet Secretary Shona Robison said, "We are committed to advancing equality for women and protecting women's rights. That commitment is not affected by our support for trans rights."[1] The draft Bill of which she spoke aims to streamline the process for those seeking to obtain a Gender Recognition Certificate (GRC) and so amend the sex on their birth certificate to the gender with which they identify.[2] Those in favour of the draft Bill contend that it improves rights recognition and reduces discrimination for trans people, including trans women. Those in opposition argue that it puts the rights and equality of cis women at risk.[3] In Scotland, as in other parts of the United Kingdom (UK), this debate is contentious and highly polarised. Those advancing the draft Bill evidence cognisance of this fact, and of the competing views, as Robison's words attest.

Given the relevance of the debate to efforts to reduce discrimination against women—trans and cis—this article's contribution is to analyse the extent of the draft Bill's adherence to the Convention on the Elimination of All Forms of Discrimination against Women (CEDAW). Conducting this analysis, I argue that the draft Bill is in adherence with CEDAW because of its ambition to reduce discrimination faced by trans women through means which in no way increase the risk of discrimination against cis women. Drawing on postmodernism, however, I will contend that it does not go far enough in this attempt because its conceptualisation of gender is socially constructionist, and thus tacitly acknowledges the role of biological determinism. As such, it renders an opportunity for anti-trans movements to challenge the draft Bill on the grounds that trans women are not "real" women and reduces the opportunity for

[1] Shona Robison, words recorded in 'Official Report Draft: Meeting of the Parliament (Hybrid)', Session 6, The Scottish Parliament (3 March 2022) 65–66.

[2] Scottish Government, 'Gender Recognition Reform Bill' (Scottish Government) <https://www.gov.scot/news/gender-recognition-reform-bill/> accessed 3 April 2022.

[3] "Cis" refers to people who live in the gender which is the same as the sex that was assigned at birth.

coalitional work between feminist and queer rights groups on the emancipation of all women.

To advance my discussion, I first review the pertinent legislation to chart the current process applicants must undertake to obtain a GRC under UK and Scots law. This groundwork enables me to draw out the key features of the proposed reforms in the draft Bill and its relevance to CEDAW. Second, I set out a conceptual framework of classical sociological, social constructionist and postmodern conceptualisations of "gender" to explore those found within the draft Bill, and in the positions of those who support and oppose its reforms. To support this work, I have conducted substantial primary qualitative analysis of public consultation responses published on the Scottish Government's website. This supports the final section of this article, Section IV, where I critically assess the adherence of the draft Bill against a framework of two of CEDAW's most relevant articles—Articles 2 and 5, those relating to the elimination of discrimination and changing social and cultural patterns—and the CEDAW Committee's General Recommendation 28 (GR28) which clarifies CEDAW's intent with regards to gender- as well as sex-based discrimination. This work, and this article's premise and argument, is underpinned by a normative commitment to protecting and strengthening trans rights in Scotland.

II. Doctrinal Analysis of Gender Recognition Legislation in Scotland

To obtain a GRC under UK and Scots law, an applicant must currently fulfil criteria set out in the Gender Recognition Act 2004. A marked step forward for the rights of trans people, this Act was introduced following two cases that went to the European Court of Human Rights in 2002. In *Christine Goodwin v the United Kingdom*, the Court found that the UK had breached the rights of Goodwin, a trans woman, under Article 8 (right to respect for private life), and Article 12 (right to marry) of the European Convention on Human Rights.[4] In *I v the United Kingdom*, the Court found the same with regard to a second trans woman.[5] Both cases reflected binary understandings of gender,[6] and in both the Court reasoned that the practice of restricting gender in national law to the sex registered at birth constituted a risk of violations to private life. This was considered to be as a result of trans people being regularly required to reveal their registered birth sex, for example in pre-employment checks. It was further reasoned to constitute a

4 *Christine Goodwin v the United Kingdom* App no 28957/95 (ECHR, 11 July 2002) para 124 (1)–(2).

5 *I v the United Kingdom* App no 25680/94 (ECHR, 11 July 2002) paras 73 and 84.

6 Ralph Sandland, 'Crossing and Not Crossing: Gender, Sexuality and Melancholy in the European Court of Human Rights' (2003) 11 Fem Leg Stud 191.

violation of their right to marriage, for English law would legally recognise their marriage only to other women, even though they were living as women.[7] Therefore, it was held that the UK must establish procedures to correct these violations.[8] The subsequently-introduced Gender Recognition Act 2004 established the right for trans people in the UK to legally change their gender - including an amendment to their birth certificate—by obtaining a GRC through three 'tracks': standard,[9] alternative,[10] and overseas.[11] The track considered in this article is the standard track, as that is the one the draft Bill seeks to streamline.[12] To successfully apply for a GRC under this track, applicants must: have been diagnosed with gender dysphoria; have lived in their "acquired gender"[13] for two years immediately prior to their application; and intend to live in their "acquired gender" for the rest of their life.[14] A second piece of legislation should be noted as constituting an important aspect of the legal framework which supports the rights of trans people. It is the Equality Act 2010, which legally protects people from discrimination in the workplace and in wider society,[15] and includes "gender reassignment" as a protected characteristic.[16] It contains specific provision on trans rights, for example prohibiting gender reassignment discrimination in access to and provision of separate and single-sex services.[17]

In a Consultation Paper shared in 2019, whereby the Scottish Government made the case for the draft Bill, former Cabinet Secretary Shirley-Anne Somerville MSP acknowledged the process to obtain a GRC under the 2004 Act is arduous, further acknowledged the discrimination trans people face in society, and noted the Scottish Government's responsibility to comply with international human rights law to reduce it.[18] Therefore, the draft Bill proposes to streamline the process of applying for a GRC in a number of ways. First, it provides that eligible applicants[19] will no longer have to provide medical reports or evidence of a medical diagnosis of gender dysphoria, with a statutory declaration on their intent and

[7] Rhona KM Smith, 'Goodwin v. United Kingdom App. No. 28957/95 and I. v. United Kingdom. App. No. 25680/94' (2003) 97(3) AJIL 659, 660-661.

[8] Betty C Burke, 'No Longer the Ugly Duckling: The European Court of Human Rights Recognizes Transsexual Civil Rights in Goodwin v. United Kingdom and Sets the Tone for Future United States Reform' (2004) 64 LA Law Rev 643, 643.

[9] Gender Recognition Act 2004, s 1(1)[a].

[10] ibid, s 3A(1)–(6).

[11] ibid, s 1(1)[b].

[12] Scottish Government, 'LGBTI and gender recognition' (Scottish Government) <https://www.gov.scot/policies/lgbti/gender-recognition/> accessed 11 April 2022.

[13] I have placed quotation marks around "acquired gender" here to reflect the exact terminology in the Act, which is notable and which I shall return to for analysis in Section II. Conceptualisations of Gender.

[14] Gender Recognition Act 2004, s 2(1)[a]–[c].

[15] Equality Act 2010, Introductory Text.

[16] ibid, ss 4 and 7.

[17] ibid, Explanatory Notes, part 16 schedule 3 part 7 para 28.

[18] Scottish Government, 'Gender Recognition Reform (Scotland) Bill: A consultation by the Scottish Government' (December 2019) 2.

[19] Those born, or habitually resident, in Scotland.

admissibility deemed sufficient.[20] Second, it reduces the period they must have lived in their "acquired gender" from two years to three months,[21] plus an additional three month reflection period.[22] Although the CEDAW is not mentioned in the document, the Yogyakarta Principles - nonbinding Principles developed by experts in 2007, which set out how the framework of international human rights law should apply to those of diverse sexual orientation and gender identity - were noted as a development beyond Scotland that suggested the process by which gender recognition before the law should be simplified.[23] In particular, it cited Principle 3c, which contends that States should take all necessary steps to ensure State-issued identity papers (including birth certificates) "reflect the person's profound self-defined gender identity".[24]

The key points to be drawn from this analysis are as follows. First, the Scottish Government deems that the current process for securing a GRC under the Gender Recognition Act 2004 is arduous and risks dissuading trans people from attempting to secure legal recognition for their gender, thus exposing many to the risk of discrimination. Second, although the CEDAW is not directly referenced anywhere in the draft Bill or the Scottish Government's justification for it, it is relevant because its object is to eliminate discrimination faced by women, including trans women. It is further relevant because the Scottish Government's use of the Yogyakarta Principles as part of their justification indicates their willingness to consider and adopt international standards to support their case. Third, language within the draft Bill—specifically use of the term "acquired gender"—is unchanged from that used in the Gender Recognition Act 2004, which in turn mirrors the binaries inferred in the *Goodwin* and *I* cases. This leads me to discuss conceptualisations of gender within the draft Bill.

III. Conceptualisations of Gender Set Out in the Draft Bill and by Consulted Parties

Historically, Western conceptualisations of gender are informed by two main schools of thought: classical sociological and social constructionist. The former, dominant in the late nineteenth and early-mid twentieth centuries, "drew on and contributed to understandings of sex, gender and sexuality as binary categories

[20] Scottish Government, 'Gender Recognition Reform (Scotland) Bill [As Introduced] 2022', s 8(C)(1)[a].
[21] ibid, s 8(C)(1)[a][iii].
[22] ibid, s 8(B)(5).
[23] Scottish Government (n 18) 18–19 paras 3.38–3.39.
[24] International Commission of Jurists (ICJ), 'Yogyakarta Principles: Principles on the application of international human rights law in relation to sexual orientation and gender identity' (March 2007) 3(c).

ordained by nature."[25] It therefore took an essentialist view of gender as equivalent to sex - with sex determined by physical characteristics - and set out binary categories such as man and woman, masculine and feminine. Challenging this view, social constructionism, emerging in the mid-late twentieth century, "[shifted] away from biologically based accounts of gender to social analysis".[26] Although not denying the role of biology in determining sex, social constructionists argued that social and cultural factors define gender, and therefore introduced the idea of the sex/gender binary. More recently, a third theoretical shift - postmodernism - has occurred that contends that gender is performatively enacted. Postmodern gender theorists such as Judith Butler argue that this means "[gender] is real only to the extent that it is performed."[27] As a result, postmodernism contends that gender is not fixed and immutable, but is co-constitutive of cultural and social subjectivities that mean it can be multiple, nonstatic and context-specific. Decolonial scholarship supports the postmodern view, with Mohanty arguing that the supposed "universality" of woman as a subject of patriarchal oppression is a discursive tool used by Western feminisms to construct the "other" of what she terms the "uber-oppressed" Third World Woman.[28] Ethnographic studies also contest notions of universality in gender, with hundreds of diverse gender identities recorded globally including the *kothi* of India[29] and the Ugandan *mudoko dako*.[30] In the West, a number of "gender categories" continue to emerge including intersex, agender, and gender questioning. These categories align with postmodernism in transcending the sex/gender binary, and map to a new form of identity politics which can be a powerful social movement promoting the development and enjoyment of rights for groups within particular categories.

This has given some pause for thought, however. Mohanty raises the concern that these identities are new "boxes" that risk exclusion of those who do not neatly fit and threaten the potential formation of solidarities between groups. She argues instead for a recommitment to "complex politics of antiracist, anti-imperialist feminisms" which defy neat categorisation.[31] Postmodern gender theorists tend to agree with the caution in this assessment; consider Otto, who argues for "more feminist and queer coalitional work and the adoption of a

25 Diane Richardson, 'Conceptualising Gender' in Diane Richardson and Victoria Robinson (eds), *Introducing Gender and Women's Studies* (London: Palgrave Macmillan 2008) 4.

26 ibid 5.

27 Judith Butler, 'Performative Acts and Gender Constitution: An Essay in Phenomenology and Feminist Theory' (1988) 40(3) Theatre J 519, 527.

28 Chandra Talpade Mohanty, 'Under Western Eyes: Feminist Scholarship and Colonial Discourses' (1984) 12(3) Bound 2 333, 334.

29 Soumi Dey, 'Being A 'Kothi': An Ethnographic Interrogation with A Male Transgender in Kolkata, India' (2013) 11(6) IOSR-JHSS 51.

30 Sylvia Tamale, 'Out of the Closet: Unveiling Sexuality Discourses in Uganda' in Catherine M Cole et al (eds), *Africa After Gender?* (Indiana University Press 2007) 17, 18.

31 Chandra Talpade Mohanty, 'Transnational Feminist Crossings: On Neoliberalism and Radical Critique' (2013) 38 Signs (Chic) 967, 987.

performative understanding of 'sex'"[32] and Butler, who contends that feminists should resist the "presumed universality and unity of the subject of feminism" (woman), to defy juridical knowledge structures which have created and reinforced this category and the patriarchal oppression it engenders.[33] This, as my analysis will highlight, is particularly relevant when it comes to the rights of trans women and those of cis women.

Using the preceding theoretical framework, the conceptualisations of "gender" in the draft Bill can now be analysed. The draft Bill refers to "either gender",[34] and to "acquired gender", in relation to "the gender in which the [applicant for a GRC] is living when the application is made."[35] "Either" suggests a binary conceptualisation of gender, but reference to "in which the person is living" also reflects an acknowledgement that gender is in part determined by social factors and not wholly defined according to sex. The term "acquired gender" —mirroring that found in the Gender Recognition Act 2004—indicates a view that gender can be changed or "acquired" according to factors which may or may not include social and cultural determinants. Therefore, the draft Bill's conceptualisation of gender leans towards the social constructionist view and does not deny the role of either sex (biologically determined) or gender (socially and culturally influenced). In this way its conceptualisation of gender is similar to that set out in the Yogyakarta Principles, which determine that gender identity refers to individuals' "deeply felt internal and individual experience of gender, which may or may not correspond with the sex assigned at birth".[36] Otto, however, registers concern that such conceptualisations "step away from hard-won social constructivism",[37] relying closely as they do—despite acknowledging the role culture can play—on a continued understanding of bio-determinism. In this way, and much as Sandland argues regarding the *Goodwin* and *I* cases,[38] the draft Bill is conservative in shoring up traditional binary ideas of gender and sex.

This point becomes more important when considered against the different conceptualisations of gender evident among those consulted on the draft Bill's reforms. The Scottish Government held two public consultations on the changes proposed and published an analysis of the over 16,000 responses it received to the second of these, which was held between December 2019 and March 2020. This analysis determines that opinions on the draft Bill fell into two main "camps": those broadly in support of the proposed changes (largely comprised of Lesbian,

32 Dianne Otto, 'Queering gender [identity] in international law' (2015) 33(4) Nordic Journal of Human Rights 299, 299.

33 Judith Butler, *Gender Trouble: Feminism and the Subversion of Identity* (2nd edn, Routledge Classics 2006) 5.

34 Scottish Government (n 20) s 8(A)(1).

35 ibid, s 8(C)(3).

36 International Commission of Jurists (ICJ) (n 24) preamble 8.

37 Otto (n 32) 301.

38 Sandland (n 6) 191.

Gay, Bi, and Trans (LGBT+[39]) groups, youth groups, local authorities, and third sector organisations),[40] and those broadly opposed (mainly "Women's Groups and Religious or Belief Bodies").[41] Over 200 organisations submitted responses, which have been published, and primary qualitative analysis thereof—conducted by this author through extensive desk-based research—reveals trends in the conceptualisations of gender within the camps. The former camp was broadly aligned to the social constructionist view, acknowledging the distinction between sex and gender, and expressing support for the view that trans people should be recognised according to their gender, not their sex characteristics. In addition, many went further than social constructionism towards postmodern conceptualisations, acknowledging that gender goes beyond the binary by noting concern that the draft Bill contains no provision for non-binary people. This view was expressed not just by LGBT+ and queer organisations - such as Stonewall Scotland,[42] Argyll & Bute Trans Youth,[43] and Beyond Gender[44]—but by local authorities and workers' unions such as Aberdeenshire Council[45] and UNISON.[46] Conversely, the second camp was much more aligned to the classical sociological view. Religious groups were particularly vehement on the view that sex is biologically determined. Catholic Truth stated: "There is absolutely no scientific or medical evidence to support the belief that a man can become a woman and a woman can become a man".[47] Many women's groups too noted disbelief that trans women are really women, therefore revealing an essentialist view of the biological determinism of sex and gender by implication. For example, Fife Women's Aid noted that the draft Bill "fail[s] to assess the impact on women who require single-sex or sex-segregated services and those who require care … from workers who

39 The "+" is intended to encompass other gender- and sex-nonconforming identities including asexual, queer and intersex.

40 Scottish Government, 'Gender Recognition Reform (Scotland) Bill: Analysis of responses to the public consultation exercise' (September 2021) 5 para 1.22.

41 ibid 5 para 1.24.

42 Stonewall Scotland, 'Response' (Scottish Government, 2 September 2021) <https://www.gov.scot/collections/gender-recognition-reform-scotland-bill-consultation/#consultationresponses> accessed 11 April 2022 10.

43 Argyll & Bute Trans Youth, 'Response' (Scottish Government, 2 September 2021) <https://www.gov.scot/collections/gender-recognition-reform-scotland-bill-consultation/#consultationresponses> accessed 11 April 2022 2.

44 Beyond Gender, 'Response' (Scottish Government, 2 September 2021) <https://www.gov.scot/collections/gender-recognition-reform-scotland-bill-consultation/#consultationresponses> accessed 11 April 2022 2.

45 Aberdeenshire Council 'Response' (Scottish Government, 2 September 2021) <https://www.gov.scot/collections/gender-recognition-reform-scotland-bill-consultation/#consultationresponses> accessed 11 April 2022 1.

46 UNISON SCOTLAND, 'Response' (Scottish Government, 2 September 2021) <https://www.gov.scot/collections/gender-recognition-reform-scotland-bill-consultation/#consultationresponses> accessed 11 April 2022 2.

47 Catholic Truth, 'Response' (Scottish Government, 2 September 2021) <https://www.gov.scot/collections/gender-recognition-reform-scotland-bill-consultation/#consultationresponses> accessed 11 April 2022 1–2.

are female”, thus implying that trans women aren’t female.[48] Meanwhile, Portobello Against Misogyny argued, “[s]ex is clearly defined in law, and women, as a group, have fought for and won rights and protections on the basis of sex.”[49] This final point is key: it sets up the debate that, by supporting the rights of trans women and seeking to reduce the discrimination they face, the Scottish Government risks increasing the discrimination faced by cis women, won on the basis of sex. I now consider this debate. Before doing so, it is important to note that not *all* women’s groups opposed the draft Bill. Engender[50] (a feminist organisation working to realise women’s equality in Scotland) and Wise Women Glasgow[51] (which works to support women with personal safety concerns) were among those in broad support of its reforms.

IV. Critical Analysis of the Draft Bill’s Adherence to CEDAW

CEDAW was adopted by the United Nations (UN) General Assembly on the 18 December, 1979, with the progress States Parties make in their obligations to it monitored by the CEDAW Committee. The UK is a State Party, having ratified the Convention in 1986,[52] in so doing making no reservations related to Articles 2 or 5.[53] Although Scotland, as a member of the Union, cannot ratify the Convention directly, the Scotland Act 1998 designates the observation and implementation of international obligations as nonreserved matters, which means the Scottish Parliament is required to legislate to give effect to those obligations[54]—including CEDAW. To critically analyse the Scottish Government’s draft Bill for its adherence to CEDAW, I have selected for close examination CEDAW’s Articles 2

48 Fife Women’s Aid, ‘Response’ (Scottish Government, 2 September 2021) <https://www.gov.scot/collections/gender-recognition-reform-scotland-bill-consultation/#consultationresponses> accessed 11 April 2022 11.

49 Portobello Against Misogyny, ‘Response’ (Scottish Government, 2 September 2021) <https://www.gov.scot/collections/gender-recognition-reform-scotland-bill-consultation/#consultationresponses> accessed 11 April 2022 1.

50 Engender, ‘Response’ (2 September 2021) (Scottish Government, 2 September 2021) <https://www.gov.scot/collections/gender-recognition-reform-scotland-bill-consultation/#consultationresponses> accessed 11 April 2022 1.

51 Wise Women Glasgow, ‘Response’ (Scottish Government, 2 September 2021) <https://www.gov.scot/collections/gender-recognition-reform-scotland-bill-consultation/#consultationresponses> accessed 11 April 2022 3.

52 UN OHCHR, ‘Ratification Status for United Kingdom of Great Britain and Northern Ireland’ (UN OHCHR) <https://tbinternet.ohchr.org/_layouts/15/TreatyBodyExternal/Treaty.aspx?CountryID=185> accessed 3 April 2022.

53 United Nations, ‘Declarations, reservations, objections and notifications of withdrawal of reservations relating to the Convention on the Elimination of All Forms of Discrimination against Women’, CEDAW/SP/2006/2 (10 April 2006) 31–33.

54 UK Government, ‘Scotland Act 1998 Explanatory Notes’ (UK Government) <https://www.legislation.gov.uk/ukpga/1998/46/notes/division/5/5/9/3?view=plain> accessed 3 April 2022 para 7(2)[a].

(eliminating discrimination) and 5 (modifying social and cultural patterns). I also consider the CEDAW Committee's GR28, which clarifies the scope and meaning of the Convention's Article 2 with regards to gender- as well as sex-based discrimination.

A. Article 2

Article 2 of CEDAW requires that "States Parties condemn discrimination against women in all its forms, [and] agree to pursue by all appropriate means and without delay".[55] It further sets out that States Parties should do so by embodying the principle of equality in legislation;[56] refraining from engaging in any act or practice of discrimination against women;[57] and taking all appropriate measures to modify or abolish existing laws that discriminate against women.[58] I argue that the draft Bill complies with CEDAW on this point because it seeks to reduce discrimination faced by trans women in their recognition before the law, and poses no discriminatory threat to cis women. In addition, although it does not completely remove the discrimination faced by trans women regarding their legal gender, it evidences the Scottish Government at least making progress in this regard. To develop this argument, I first determine the types of discrimination faced by trans women in Scotland which result from, or are correlated to, challenges they face in having their gender recognised before the law. I then determine how the draft Bill seeks to address this, before elucidating the two main critiques of the draft Bill with regard to discrimination elimination.

The CEDAW Committee highlighted discrimination faced by trans women as a concern in their concluding observations on the UK's eighth periodic report in 2019. Specifically, the Committee called on the UK - and ergo Scotland - to, "Review and amend the public sector equality duty in order to address situations of intersectional forms of discrimination, such as discrimination faced by … transgender women."[59] Although there is no precise data on the number of trans women in Scotland, a needs assessment in 2018 estimated the population of trans people in Scotland to number just under 24,000.[60] It can therefore be inferred that trans women compose a significant minority. There is a significant body of evidence to suggest that they experience discrimination. 41% of trans people

55 UN OHCHR, 'Convention on the Elimination of All Forms of Discrimination against Women 1979 [CEDAW]' (1979) art 2.

56 ibid, art 2(a)–(b).

57 ibid, art 2(c).

58 ibid, art 2(f).

59 UN Committee on the Elimination of Discrimination against Women, 'Concluding observations on the eighth periodic report of the United Kingdom of Great Britain and Northern Ireland', CEDAW/C/GBR/CO/8 (8 March 2019) para 15(c).

60 Rachel Thomson, Jessica Baker and Julie Arnot, 'Health Care Needs Assessment of Gender Identity Services', Scottish Public Health Network (ScotPHN) (May 2018) 11.

responding to a 2018 survey by Stonewall, the UK's largest LGBT+ rights organisation, said they had experienced a hate crime or incident because of their gender identity in the last twelve months.[61] Given the current challenges associated with obtaining a GRC, many trans people do not have one. These people may subsequently be forced into revealing the sex they were assigned at birth in pre-employment checks, meaning workplace discrimination is particularly relevant. A 2021 survey conducted by LGBT Health and Wellbeing, a charity working in support of trans rights in Scotland, found that trans people experience discrimination at work and when looking for work, with 40% of respondents saying their trans identity had a "quite" or "very" negative impact on their job prospects.[62] 34% of respondents identified as "woman" or "transwoman".[63]

The draft Bill reduces discrimination potential of the sort elucidated in the preceding paragraph in the following ways. First, because it seeks to make it easier for trans people to obtain a GRC, it is likely that challenges faced in the workplace will be reduced; because more trans people will have GRCs, they will no longer have to reveal the sex they were assigned at birth. Second, it alleviates requirements of proof that applicants have lived in their "acquired gender" for two years, which reduces the likelihood of discrimination against particularly vulnerable trans people, like homeless women, who may not have documentation like driver's licences which constitute evidence. Third, it reduces the risk of circularity whereby employers seek a GRC as a condition of change of name at work, in turn reducing trans people's chances of successfully applying for a GRC because they cannot use evidence from work to support their claim. Though illegal, this practice is noted to be common.[64] Finally, the draft Bill is in adherence with CEDAW Article 2(f) which notes States Parties should, "modify or abolish existing laws … which constitute discrimination against women".[65] This is because it changes the legislative environment in favour of trans people, making it easier for them to get a GRC and thus benefit from the provisions for gender recognition set out in the Gender Recognition Act 2004.

There are two main critiques of the draft Bill that indicate potential non-adherence to Article 2. The first comes from those who broadly support the proposed reforms. Their critique is that the requirement for an applicant to live in their acquired gender for three months is arbitrary and constitutes a risk of discrimination. Although it represents a reduction of the current requirement to live in the acquired gender for twenty-four months, it still puts a burden of proof upon the applicant. This, I concur, has grounds and suggests the draft Bill does

61 Government Equalities Office, 'Trans People in the UK' (2018) 1.

62 LGBT Health & Wellbeing, 'Trans People and Work: Survey Report' (2021) 12 and 5 respectively.

63 ibid 9.

64 UNISON SCOTLAND (n 47) 1.

65 UN OHCHR (n 56) art 2(f).

not go as far as it might to reduce the potential for discrimination against trans women. I return to this point in my discussion of Article 5.

I argue, however, that the second—more widely touted—critique has no grounds. It comes from those who broadly oppose the proposed changes on the grounds that the removal of the requirement for a diagnosis of gender dysphoria will make the system open to abuse "allowing predatory men to access women's safe spaces".[66] Here it can be inferred that "women" means "cis women". There is little evidence to support these concerns. Former Cabinet Secretary Shirley-Anne Somerville, introducing the second public consultation, was clear that risks to women's spaces are not posed by trans women, but cis men. She noted: "[The concerns] are about men who seek to abuse women … That's not a … problem created by, or the fault of, trans people."[67] Her assertion is backed up: as my doctrinal analysis highlighted, the Equality Act 2010 permitted trans people equal access to women-only spaces, and in the intervening twelve years there is no indication that attacks against women in said spaces have increased.[68] It is also notable that the rape support centres which responded to the second consultation broadly supported the proposed reforms. The Edinburgh Rape Crisis Centre noted, "The changes proposed in this Bill will have no negative impact on our ability to support the survivors of rape and sexual assault" and argued that they will likely make it easier for trans people to seek support.[69] The Forth Valley Rape Crisis Centre agreed with this assessment.[70] If there were any real risks to women's safety, they would almost certainly have opposed the reforms. In sum, it is this article's assessment that the draft Bill adheres with CEDAW's Article 2 and is indicative of the Scottish Government making efforts to improve the legislative environment for trans women, per Article 2(f) of CEDAW. Furthermore, there is no evidence to suggest that its proposed reforms increase the risk of discrimination against cis women.

B. General Recommendation No 28

GR28 aims to clarify the scope and meaning of Article 2 of the Convention, specifically clarifying that *gender*-based as well as *sex*-based discrimination should

66 Scottish Government (n 41) iii.

67 Scottish Government (n 18) 29 para 5.06.

68 Mermaids, 'Safety & Dignity: trans rights are no threat to single-sex spaces' (Mermaids, 14 June 2020) <https://mermaidsuk.org.uk/news/safety-and-dignity/> accessed 4 April 2022.

69 Edinburgh Rape Crisis Centre, 'Response' (Scottish Government, 2 September 2021) 3 <https://www.gov.scot/collections/gender-recognition-reform-scotland-bill-consultation/#consultationresponses> accessed 11 April 2022.

70 Forth Valley Rape Crisis Centre, 'Response' (Scottish Government, 2 September 2021) 3 <https://www.gov.scot/collections/gender-recognition-reform-scotland-bill-consultation/#consultationresponses> accessed 11 April 2022.

be tackled.[71] Although not part of the Convention's original text, it is an important element of the CEDAW framework and reflects the dynamism with which the CEDAW Committee interprets the Convention text.[72] Significantly, it evidences a development in CEDAW's conceptualisation of gender, defining "gender" as "socially constructed identities, attributes and roles for women and men and society's social and cultural meaning for these biological differences resulting in hierarchical relationships between women and men".[73] Therefore, it evidences a definition of "woman" that is not solely based on sex characteristics, and—as Meyer contends—is therefore inclusive of trans women.[74] This article has argued that the draft Bill conceptualises gender in a way that understands the social and cultural meanings attached to biological differences. Correspondingly, it seeks to make it easier for trans people to obtain GRCs by diminishing the role of sex-based characteristics in influencing gender recognition. As such, the draft Bill adheres to GR28.

I also contend, however, that neither GR28 nor the draft Bill go as far as they might to eliminate discrimination against trans women. This is because the conceptualisations of gender they contain continue to reflect biological determinism. GR28 defines "sex" as "refer[ring] to biological differences between men and women"[75] and, as previously discussed, the draft Bill also tacitly acknowledges the role of sex-based differences in determining an individual's gender. As such, both reaffirm social constructionist conceptualisations of gender which offer an easy argument for those who oppose the draft Bill: namely, that trans women are not really women because their sex characteristics mean they are "men". In tacitly reinforcing this categorisation, the draft Bill is active in the production of a category "woman" which is not as inclusive as it might be. It thus sets up the potential for opposition between women and queer rights groups, who should rather act in coalition to contend patriarchal oppression. Otto makes the point that biological determinism orders women's treatment in international law, and reflects "men" and "male" as the full standard of humanity to which women and other genders must measure up.[76] She therefore contends that feminists and queer rights activists must form coalitions to challenge this determinism to "develop a more liberatory and inclusive conception of gender in international

[71] UN Committee on the Elimination of Discrimination Against Women, 'General recommendation No. 28 on the core obligations of States parties under article 2 on the Convention on the Elimination of All Forms of Discrimination against Women', CEDAW/C/GC/28 (16 December 2010) para 5.

[72] Rikki Holtmaat, 'The CEDAW: a holistic approach to women's equality and freedom' in Anne Hellum and Henriette Sinding-Assen (eds), *Women's Human Rights: CEDAW in International, Regional and National Law* (Cambridge 2013) 95, 104.

[73] UN Committee on the Elimination of Discrimination Against Women (n 72) para 5.

[74] Elise Meyer, 'Designing Women: The Definition of Women in the Convention on the Elimination of All Forms of Discrimination against Women' (2016) 16(2) Chi J Int'l L 553, 556.

[75] UN Committee on the Elimination of Discrimination Against Women (n 72) para 5.

[76] Otto (n 32) 302.

(and domestic) law"[77] which will support both women's rights and those of people of diverse gender identity. This, as my preceding analysis has shown, has not been the case with regards to the draft Bill, which some (cis) women's groups perceive as a threat to their rights and autonomy, and thus contend vigorously. I conclude this section by reaffirming that although the draft Bill is in adherence with CEDAW's GR28, this is in part because of its acknowledgement of the role biological determinism plays in determining the category "woman", thus shoring up division between queer and women's rights groups who should rather be in coalition for the active emancipation of all women. As such, it does not go as far as it might to reduce discrimination faced by trans women; nor, in fact, does CEDAW.

C. Article 5

Article 5 requires States Parties to take all appropriate measures to "modify the social and cultural patterns of conduct of men and women."[78] Its intent is to eliminate prejudices and practices that are based on the idea of the inferiority or superiority of either of the sexes,[79] or on stereotyped gender roles for men and women. It is relevant here because its paragraph (a) acknowledges the role that society and culture play in the creation and performance of gender. This article is therefore "the least gender-specific provision in CEDAW"[80] and, although it does not acknowledge the existence of sexes or genders in addition to men and women, it can be read as "putting an obligation on States Parties to combat systemic or structural gender discrimination."[81] In this obligation, strong political will to create structural change is critical. It is therefore of note that the Scottish Government conducted two of its largest ever public consultations[82] on the changes proposed by the draft Bill, as it suggests that it conceives of these changes as constituting cultural change. Further, by seeking to make the changes inclusive and consultative, it has aimed to create an enabling environment for debate on the key issues to facilitate said change. This is significant because the CEDAW Committee has previously recommended that States Parties "intensify cooperation … with civil society organisations, women's groups and community leaders, traditional and religious leaders" in seeking to enact cultural change for the elimination of

[77] ibid 300.
[78] UN OHCHR (n 56) art 5(a).
[79] The author acknowledges that it is simplistic to refer to "either" sex, as there are sex characteristics and sexual identities which do not fit this binary (for example, intersex).
[80] Gabrielle Simm, 'Queering CEDAW? Sexual Orientation, Gender Identity and Expression and Sex Characteristics (SOGIESC) in International Human Rights Law' (2020) 29(3) Griffith Law Review 374, 385.
[81] Holtmaat (n 73) 96.
[82] Robison (n 2) 62.

discrimination against women,[83] an entreaty it makes often.[84] Therefore, I contend that the draft Bill—and the process through which the Scottish Government has gone in consulting on its provisions—is in adherence with Article 5.

I add, however, that once again it does not go as far as it might because of its retention of the requirement that someone seeking to obtain a GRC must live in their acquired gender prior to making an application. Although the draft Bill significantly reduces this period from twenty-four months to three months, this nonetheless raises important questions around what constitutes proof of living in an acquired gender. The requirement is demonstrative of a regressive understanding that "living in" a particular gender requires acting according to the traditional and societal gender roles that Article 5 obligates States Parties to challenge. In addition, although the Scottish Government has noted concern that the tone of debate on trans rights is polarised,[85] it has not helped this by characterising one of the main parties opposed to the draft Bill to be women's rights groups, which it does in analysis of the second public consultation.[86] As my own analysis has shown, a number of women's rights were actually in *favour* of the proposed reforms. To ignore this fact is to further polarise debate, falsely pitting the rights of trans women against those of cis women.

V. Conclusion

In this article, I have critically analysed the extent to which the draft Bill adheres to CEDAW and found that it is in adherence with the framework of Article 2, Article 5, and GR28 against which I have examined it. It aligns with Article 2 in seeking to reduce the discrimination faced by trans women, and—contrary to the argument of those who allege it constitutes a risk to the rights of cis women—constitutes no such risk. It further adheres to GR28 as it seeks to reduce *gender*-based discrimination, and with Article 5 because it—and the process through which the Scottish Government have gone to consult on its changes—evidences strong political will to facilitate social and structural change with regards to gender. In these ways, it demonstrates a significant improvement on the provisions contained within the Gender Recognition Act 2004 with regard to the process of obtaining a GRC. The implications of this are positive for the advancement of trans rights; the CEDAW has significant normative power in influencing national legislature, and that the draft Bill adheres to its provisions clearly enhances its legitimacy under Scots and UK law.

[83] UN Committee on the Elimination of Discrimination against Women, 'Concluding observations on the Committee on the Elimination of Discrimination against Women: Nigeria', CEDAW/C/NGA/CO/6 (8 July 2008) para 323.

[84] Holtmaat (n 73) 121.

[85] Robison (n 2) 63.

[86] Scottish Government (n 41) 5 para 1.24.

This article, however, has further argued that the draft Bill does not push and challenge the CEDAW as far as it might. This damages the realisation of trans women's rights and limits the draft Bill's potential for the facilitation of queer and feminist coalitional work to reduce patriarchal structural oppression. Indeed, the draft Bill reflects a conceptualisation of gender that is aligned with social constructionist views, therefore continuing to acknowledge biological determinism and to reinforce oppressive structures within society which themselves rely on biological determinism and the order of the sexes. Furthermore, the Scottish Government's summary of the second public consultation's outcomes has in some ways further polarised the debate on trans rights. This is because it suggests that women's rights groups are generally in opposition to the draft Bill's proposed reforms, which—as my analysis has shown—is not the case. The implication of this is to further polarise the debate on trans rights, making an already fractious public debate even more so. In turn, this creates the risk of increased discrimination against trans women who are perceived to be men seeking to threaten the safety and security of cis women.

Therefore, this article concludes by contending that both the draft Bill and the CEDAW could do more for the elimination of discrimination against all women—cis and trans—by adopting a postmodern conceptualisation of gender which defies categorisation of oppressed subjects in legislative frameworks and focusses clearly on tackling gendered inequalities for the active emancipation of all women.

Cambridge Law Review (2022) Vol VII, Issue 2, 72–102

An Assessment of the Effectiveness of the Unfair Prejudice Remedy in UK Company Law: How can we Guarantee Appropriate Judicial Discretion?

ZIYUAN LI*

ABSTRACT

In the UK, members of a company can petition the court for a remedy in respect of conduct by other members that unfairly prejudices their interests under section 994 of the Companies Act 2006. Indeed, the breadth of interpretive judicial discretion concerning the core wording of s 994 (for example, the reference to 'unfairly prejudicial' and 'interest') determines the extent to which the section can act as a shield for shareholders. Since minority shareholders are vulnerable to oppression by the majority in private companies, the courts tend to show a pro-minority attitude when hearing unfair prejudice cases. Therefore, the s 994 petitions are popular with the minority shareholders. Notably, while the court's open-ended interpretation of s 994 provides a reliable safeguard for the minority shareholders' interests, it may indirectly encourage their opportunistic behaviour of abusing unfair prejudice actions. In practice, the rapidly growing number of s 994 petitions have led to this type of proceeding becoming more burdensome, thereby increasing the financial and time burden on both the petitioner and the court. Moreover, the expansive discretion has resulted in an overlap in jurisdiction between s 994 petitions, which traditionally represent personal relief, and derivative claims, which represent corporate relief. This probably opens the floodgates for minority shareholders to bring malicious claims to interfere with the affairs of the company. In this sense, the unfair prejudice remedy regime may run counter to the objectives of 'efficiency' and 'fairness' in the area of shareholder

* LLM (University of Bristol). This article corresponds to the dissertation presented for the degree of LLM, supervised by Dr Basil Salman. I am grateful to him and the anonymous reviewers for their helpful comments and feedback on earlier drafts. Any errors that remain are my own. E-mail address: lzy604109768@163.com

remedies law. Consequently, this article will attempt to explore the promising direction for improving the effectiveness of the s 994 petitions. Taking into account the legislative basis of the section, a guiding framework on the construction of appropriate judicial discretion will be proposed to better balance shareholder protection and corporate autonomy.

Keywords: unfair prejudice; interest; efficiency; fairness; judicial discretion

I. INTRODUCTION

Minority shareholder remedies are one of the hottest topics in UK company law, as a robust minority shareholder protection regime helps to build investors' confidence in their companies and the overall stock market, thus creating investment incentives.[1] In particular, the unfair prejudice remedy regime under section 994 of the Companies Act 2006 (CA 2006)[2] has been subject to considerable academic scrutiny due to its frequent use. In reality, the main target of protection under this legislation is the minority in private companies.[3] At present, the vast majority of companies registered under company law in the UK are private companies (also commonly referred to as small businesses).[4] It can therefore be argued that the unfair prejudice remedy plays an essential role in the area of shareholder remedies law in the UK.

Courts examining s 994 petitions are often mindful of the mixed commercial and personal attributes of private companies. At the inception of a private company, there is generally a tacit arrangement among the members that they will not only enjoy the profits of the company in proportion to their respective shareholdings, but will also manage the company jointly as directors.[5] Nonetheless, disagreements inevitably arise during the company operation, because shareholders usually looking out only for their own interests.[6] In such circumstances, the majority shareholders tend to vote to remove the minority from the board of directors in order to eliminate dissenting voices in the management

1 Law Commission, 'Shareholder Remedies Consultation' (1996) *Law Com No 142*, para 1.13 < https://s3-eu-west-2.amazonaws.com/lawcom-prod-storage-11jsxou24uy7q/uploads/2015/03/cp142_Shareholder_Remedies_Consultation.pdf> accessed 1 September 2021.

2 Companies Act 2006, s 994.

3 Law Commission, (n 1) para 14.5.

4 Department for Business, Energy and Industrial Strategy, 'Business population estimates for the UK and regions 2020: statistical release' (8 October 2020) <https://www.gov.uk/government/statistics/business-population-estimates-2020/business-population-estimates-for-the-uk-and-regions-2020-statistical-release-html> accessed 27 August 2021.

5 MA Iqbal, 'The Effectiveness of Shareholder Dispute Resolution in Private Companies under UK Companies Legislation: An Evaluation' (*PhD, Nottingham Trent University* 2008), 32-33.

6 DD Prentice, 'Protecting Minority Shareholders' Interests' in D. Feldman and F. Meisel (eds), Corporate and Commercial Law: Modern Developments (*Informa UK Ltd,* 1996), 80.

of the company.[7] Furthermore, due to the illiquidity of the share capital of private companies, the minority shareholders cannot easily exit the company to recover their investment.[8] Understandably, without limiting the absolute control of the company of the majority shareholders, they are very likely to flex their muscles for their self-interests at the expense of the minority.[9] In this regard, the s 994 petition can highlight its value in maintaining the delicate balance between the legitimate business decisions of the majority shareholders and the reasonable interests of the minority shareholders.

Given the sympathy for vulnerable groups, the wording of s. 994, such as the wording of 'unfairly prejudicial', has been designed to be very extensive to provide greater protection for minority shareholders. This has set a foundation for judicial practice in the exercise of the court's broad interpretive discretion.[10] However, unfettered judicial discretion may turn the umbrella of the minority shareholders into a tool that shakes the rightful dominance of the majority shareholders and interferes with corporate autonomy.[11] Such a trend has been confirmed in the case law. Firstly, the court's expansive interpretation of s 994 has encouraged the minority shareholders to submit wide-ranging and complex factual material at the pleading stage,[12] which has increased the time and cost of s 994 proceedings. Secondly, this approach to interpretation has allowed for some extension of the application of s 994 petitions from traditional personal to corporate remedies.[13] That is to say, there has been an overlap of jurisdiction between s 994 claims and derivative claims and hence a potential increased risk of abuse of s 994 petitions. Accordingly, the English Law Commission discussed these difficulties in its latest report[14] on shareholder remedies. Nevertheless, some of the relevant recommendations (such as the two statutory presumptions) made by the Law Commission were not adopted to improve the efficiency of s 994 proceedings. Nor did these proposals address how to mitigate the problems created by the overlap between s 994 and derivative claims. In this way, unfair prejudice remedies still have a long way to go before they become good laws.

[7] AD Spratlin Jr, 'Modern Remedies for Oppression in the Closely Held Corporation' (1990) 60 *Mississippi Law Journal* 405, 406-408.

[8] N Flourentzou, 'Minority Shareholders: Applicability of Unfair Prejudice' (*Shambartas*, 2014) <http://www.mslawyers.eu/images/publication_documents/Minority_Shareholders-_Applicability_of_Unfair_Prejudice.pdf > accessed 2 July 2021.

[9] A Hicks and SH Goo, *Cases and Materials on Company Law* (8th edn, Oxford University Press 2008), 425.

[10] PI Roberts and J Poole, 'Shareholder Remedies - Efficient Litigation and the Unfair Prejudice Remedy' (1999) 2 *Journal of Business Law* 38, 41-45.

[11] J Mukwiri, 'Using s.459 as an Instrument of Oppression?' (2004) 25 Company Lawyer 282, 282-284.

[12] Law Commission, (n 1) paras 1-11.

[13] Law Commission, (n 1) paras 2.1-2.26.

[14] Law Commission, 'Shareholder Remedies Report' (1996) Law Com No 246 <https://s3-eu-west-2.amazonaws.com/lawcom-prod-storage-11jsxou24uy7q/uploads/2015/03/lc246_Shareholder_Remedies.pdf> accessed 15 September 2021.

Obviously, there are two main routes to reform the unfair prejudice remedy regime, namely to amend the language of s 994 or to adjust the judicial discretion regarding s 994. However, the Law Commission persuasively argues against the former, as it would significantly limit the scope of remedies available to the minority.[15] From this logic, this article will focus on constructing a feasible judicial discretion framework to help achieve a balance between the flexibility and certainty of s 994 petitions. Under this framework, an effective unfair prejudice remedy would be the result of efficiency and fairness after considering the interests of all parties. To achieve this research objective, this article will be divided into four main sections. Following the introduction, Section II will describe the theoretical underpinnings and legislative background of the unfair prejudice remedy to explain what role the remedy needs to play in commercial life, or rather, what standards legislators expect a truly effective remedy to meet. Then, Section III will examine in detail the statutory framework of the unfair prejudice remedy system, which will give a clear picture of how much room there is for the courts' interpretive discretion to be exercised. This section will also reflect the fact that the courts' discretion is broad enough to cover most oppression of minority shareholders and to grant them appropriate remedies. Arguably, the s 994 petition is successful in terms of protecting minority shareholders. After that, Section IV will critically analyse the undesirable consequences of overly wide judicial discretion, such as the length of s 994 proceedings and the jurisdictional intersection of s 994 with derivative claims. Finally, Section V will propose an authoritative framework for guaranteeing appropriate discretion to address the above-mentioned currently unresolved difficulties. Basically, two approaches are included within the framework. The first focuses on boosting efficiency - agreeing to the statutory presumption approach proposed by the Law Commission. This article will demonstrate the feasibility of this reform measure, which was once criticised and not approved. The second approach aims to clarify the blurred line in practice between s 994 claims and derivative claims to ensure that s 994 petitions are fair to all parties.

15 Law Commission, (n 14) para 4.3-4.13.

II. The Legislative Context and Objectives of the Unfair Prejudice Remedy

A. Background: shareholder oppression in private companies

In general, the phenomenon of 'unfair prejudice' is triggered by competing positions between the majority and minority shareholders, and such conflicts are more intense in the context of private companies.

Primarily, economic theory can be used as a starting point for discussing the relationship among members of the company. In business practice, to maximise individual welfare, rational people tend to allocate their limited resources to those who can add greater value to the utility of that resource through the mutually beneficial exchange.[16] Thus, it has been maintained that the company can be understood as being seen as the nexus of a series of contractual relationships.[17] The contracting parties (for example, the shareholders and directors of the company) agree on how to distribute the profits invested in the business in accordance with the contractual arrangements.[18] Nonetheless, the majority rule,[19] the internal governance mechanism of the company, sets the stage for conflicts between the majority shareholders and the minority shareholders. Specifically, as providers of equity capital to the company, shareholders have the right to vote on important corporate matters, such as the appointment and removal of directors[20] and the approval of major corporate transactions.[21] Compared with the minority shareholders, the majority shareholders hold a controlling stake, which means that they can determine the ultimate direction of corporate decisions.[22] As opportunism encourages people to seek as much welfare as possible for themselves in business activities, majority shareholders may abuse their dominant position to 'squeeze out' minority shareholders.[23] This

16 H Atwal, 'Self-Interest, Justice and Reciprocity in Unfair Prejudice' (2004) 2004 UCL Jurisprudence Review 270, 272.

17 CRT O'Kelley, 'Coase, Knight, and the Nexus-of-Contracts Theory of the Firm: A Reflection on Reification, Reality, and the Corporation as Entrepreneur Surrogate' (2011) 35 *Seattle University Law Review* 1247, 1247-1248.

18 J Parkinson, 'Models of the Company and the Employment Relationship' (2003) 41 *British Journal of Industrial Relations* 481, 485.

19 *Prudential Assurance Co Ltd v Newman Industries Ltd (No 2)* [1982] 1 All ER 354 at 357-358.

20 B Hannigan, *Company Law* (5th edn, Oxford University Press 2018), ch 17, 439-440; Companies Act 2006, s 168.

21 ibid s 190.

22 C Fan, *Bringing Controlling Shareholders to Court: Standard-Based Strategies and Controlling Shareholder Opportunism* (Eleven International Publishing 2013), ch 2, 11.

23 ibid ch 1, 1-3.

phenomenon is typically described as the 'oppression' of the minority by the majority.[24]

Essentially, the degree of oppression suffered by minority shareholders depends to a large extent on the type of company. In UK, common forms of companies include private companies and public companies. In fact, minority shareholders in private companies may be weaker to oppression than those in public companies. Firstly, a private company is a business organisation where there is "a more intimate and intense relationship exists between capital and labour",[25] which indicates that shareholders probably expect to be substantially involved in the running of the company as directors or employees. In other words, the return on investment that shareholders want in a private company is not limited to money, but also the opportunity to manage the affairs of the company themselves.[26] In contrast, in a public company, the shareholder usually acts only as an independent investor, contributing neither labour nor management responsibilities to the company.[27] Consequently, the expected return on investment of minority shareholders in private companies may be more seriously threatened if the majority shareholders take oppressive actions such as excluding the minority from the company management or giving excessive remuneration to the controlling director.[28] Secondly, in the absence of a readily available stock market like that of public companies, it may be more difficult for dissatisfied minority shareholders in private companies who wish to exit voluntarily, as they are not free to sell their shares to outside investors.[29] In this case, the minority will be 'locked in' the company.[30] Considered the lack of ability of minority shareholders in private companies to effectively rescue their investments, legislators have been exploring a reliable mechanism to protect such shareholders.

B. HISTORY OF DEVELOPMENT: FROM 'OPPRESSION' TO 'UNFAIRLY PREJUDICIAL'—A GRADUAL EXPANSION OF JUDICIAL DISCRETION

To date, the English Law Commission has had three reform discussions targeting minority shareholder protection measures. Since general guidance standards could not be applied in every case, the courts are considered to be given a sufficiently wide discretion to ensure that the most appropriate relief can be

[24] DK Moll, 'Shareholder Oppression in Close Corporations: The Unanswered Question of Perspective' (2000) 53 *Vanderbilt Law Review* 749, 757.

[25] RB Thompson, 'The Shareholder's Cause of Action for Oppression' (1993) 48 *The Business Lawyer* 699, 702.

[26] JJ Chapman, 'Corporate Oppression: Structuring Judicial Discretion' (1996) 18 *Advocates' Quarterly* 170, 172.

[27] DK Moll, (n 24).

[28] P Paterson, 'A Criticism of the Contractual Approach to Unfair Prejudice' (2006) 27 *Company Lawyer* 204, 206-209.

[29] B Hannigan, (n 20) ch 19, 503–504.

[30] P Paterson, (n 28) 208–209.

granted to minority shareholders.[31] As a result, the relevant defining term in the minority shareholder relief law evolved from 'oppression' to 'unfairly prejudicial', leaving room for expansive interpretation by the courts in dealing with shareholder disputes.

Initially, minority shareholders faced with oppressive behaviour by the majority shareholder could merely apply to the court in limited circumstances for a just and equitable winding up as a remedy.[32] Nevertheless, the winding-up order was criticised as a radical approach because it would directly end the life of the company and deprive other shareholders of the opportunity to profit.[33] Against this background, in 1945, the Cohen Committee in its report emphasised the introduction of a statutory regime that would give the court the power to impose a just and equitable solution on the parties to a dispute.[34] Therefore, s 210 of the Companies Act 1948 (CA 1948) was introduced to focus this judicial discretion on the term 'oppressive'—if the affairs of the company were oppressive to some of the members, the members were entitled to apply to the court and the court might, at its discretion, grant such remedies as it thinks fit.[35]

Although s 210 of the CA 1948 pioneered the discretionary power given to the courts in reviewing shareholder oppression, the litigation threshold of the provision was considered too stringent to be fully utilised.[36] In 1962, the Jenkins Committee explained that the restrictive interpretation of the word 'oppression' was what made the application of s 210 too narrow.[37] A typical example is *Scottish Co-operative Wholesale Society Ltd v Meyer*, where Lord Simmonds interpreted the word 'oppression' only literally as "burdensome, harsh and wrongful",[38] meaning that unfair conduct that did not rise to the level of actual illegality was not protected by s 210.[39] In this sense, the Jenkins Committee recommended that the language of s 210 be amended to further cover more oppressive conduct.[40] Accordingly, s.75 of the Companies Act 1980[41] (later s 459 of the Companies Act

31 Cohen Committee, 'Report of the Committee on Company Law Amendment' (1945) Cmd 6659, para 60 <http://reports.mca.gov.in/Reports/17-Justice%20Cohen%20committee%20report%20of%20the%20committee%20on%20company%20law%20amendment,%201943.pdf> accessed 10 September 2021.

32 Insolvency Act 1986, s 122(1)(g).

33 Cohen Committee, (n 31).

34 ibid.

35 Companies Act 1948, s 210(1).

36 Only two examples of successful application: *Scottish Co-operative Wholesale Society Ltd v Meyer* [1959] AC 324; *Re H.R. Harmer Ltd* [1959] 1 WLR 62.

37 Jenkins Committee, 'Report of the Company Law Committee' (1962) *Cmnd 1749*, para 201 <https://www.takeovers.gov.au/content/Resources/other_resources/downloads/jenkins_committee_v2.pdf> accessed 1 September 2021.

38 *Scottish*, (n 36) at 342.

39 Jenkins Committee, (n 37) paras 203–212.

40 ibid, para 206.

41 Companies Act 1980, s 75.

1985[42]) replaced the term 'oppression' with the term 'unfairly prejudicial'. Clearly, 'unfairly prejudicial' is a broader and more general concept than 'oppression', creating favourable conditions for judges to interpret s 459 flexibly to meet the specific circumstances of different cases.[43]

In 1996, however, the open-ended approach to the interpretation of s 459 raised concerns in the Law Commission about the effectiveness of unfair prejudice actions.[44] Significantly, the Law Commission referred to the warning of Hoffmann J in *Re A Company (No 007623 of 1984)* - although giving the courts a wide discretion can safeguard the availability of s 459 petitions, such petitions might in turn become a device of oppressing majority shareholders if the breadth of jurisdiction was not carefully controlled.[45] That is to say, appropriate judicial discretion is likely to be beneficial in preventing the floodgates from opening in the jurisdiction of unfair prejudice remedies. Nonetheless, some of the creative reforms proposed by the Law Commission were set aside by the Company Law Reform Steering Group (CLRSG). This is because the CLRSG preferred to take a conservative approach in the area of company law requiring legal and practical certainty.[46] Hence, the CA 2006 does not make any changes to s 459. However, this does not demonstrate that the theme of reform in the three Law Commission reports - judicial discretion - is no longer worthy of attention. Instead, shaping appropriate interpretive discretion around the legislative objective of unfair prejudice relief may be a useful approach to preserve the effectiveness of the regime.

C. THE 'EFFECTIVENESS' OF THE UNFAIR PREJUDICE REMEDY: TWO GUIDING CRITERIA

The English Law Commission, in its review of the unfair prejudice remedy regime, correctly stated that the law needed to strike a balance between safeguarding the interests of minority shareholders and respecting legitimate business decisions of companies.[47] For one thing, a good remedy can boost the confidence of shareholders, particularly minority shareholders.[48] For another thing, corporate autonomy should not be subject to arbitrary judicial interference, given that experienced directors are in a better position than the courts to exercise

42 Companies Act 1985, s 459.

43 *Re Macro (Ipswich) Ltd* [1994] 2 BCLC 354, at 404.

44 Law Commission (n 1), paras 11.1–11.3.

45 *Re A Company (No 007623 of 1984)* [1986] BCLC 362, at 367.

46 Department of Trade and Industry, Final Report (2001) Vol I, para 7.41 <https://publications.parliament.uk/pa/cm200203/cmselect/cmtrdind/439/439.pdf> accessed 20 September 2021.

47 Law Commission (n 1) para 1.13.

48 ibid.

reasonable judgement on commercial matters in the best interests of the company.[49] To achieve a balance between these two competing objectives, this article agrees with Bahls - a viable unfair prejudice remedy should meet both the 'efficiency' and 'fairness' criteria.[50] Arguably, identifying the guiding criteria will help to examine how to improve the effectiveness of the S 994 petitions.

(i) Efficiency

Indeed, the 'efficiency' of shareholder remedies has been a key concern of the Law Commission.[51] 'Efficiency' generally stresses the need to minimise the overall waste of costs by the most reasonable solution without prejudice to the interests of any party.[52] In this regard, it will be necessary to reduce the administrative and transaction costs associated with the resolution of shareholder disputes.[53] These costs typically include the judicial costs of the courts, the litigation costs of the petitioner and the operating costs of the company.

Firstly, efficient remedies should free the courts from complex fact-finding or onerous assessments when resolving disputes, so that the costs of dispute resolution are commensurate with the benefits.[54] This would not only help to reduce the burden on the judicial system, but would also stop petitioners from struggling through lengthy proceedings. After all, the high cost of justice probably leads to a corresponding increase in the cost of litigation. In this way, minority shareholders who already lack bargaining power are more likely to shy away from litigation.[55] Secondly, from the perspective of the company's interests, unfair prejudice petitions must not be pursued at the expense of the value of the corporate assets.[56] Rather, there is a need to ensure that judicial intervention has minimal impact on the day-to-day operations of the company and that the company is not caught up in litigation that wastes money and time.[57]

[49] ibid, para 14.11.

[50] SC Bahls, 'Resolving Shareholder Dissension: Selection of the Appropriate Equitable Remedy' (1990) 15 *Journal of Corporation Law* 285, 318.

[51] Law Commission, (n 1) paras 14.11–14.14.

[52] SC Bahls, (n 50) 318-319.

[53] ibid, 327.

[54] HY Chiu, 'Contextualising Shareholders' Disputes - A Way to Reconceptualise Minority Shareholder Remedies' (2006) 5 *Journal of Business Law* 312, 314.

[55] SC Bahls, (n 50) 327.

[56] A Schultz, 'Finding the Right Remedy in Minority Shareholder Oppression Law: A Transnational Analysis of Solutions in Closely Held Corporations' (2017) 26 *Transnational Law and Contemporary Problems* 499, 505.

[57] Law Commission, (n 1) para 14.11.

(ii) Fairness

Another guiding standard to be followed in achieving effective shareholder relief is 'fairness', *i.e.,* protecting the reasonable expectations of all parties.[58] 'Fairness' can inject a degree of flexibility into company law, so that the pursuit of efficiency is not too rigid a rule. In general, the history and structure of a particular company may lead shareholders to reasonably expect certain outcomes in the event of a dispute.[59] This usually involves a proper understanding of the conflicting interests of the majority and minority within a company, particularly in the context of private companies.[60] As mentioned earlier, minority shareholders would reasonably expect that the majority would not hinder their participation in the management of the company. The majority shareholder also would expect that justice would not interfere with the normal business decisions of the company. More critically, assuming that there is a genuine need for the court to intervene in the affairs of the company, such jurisdiction should be exercised with caution so that it is not, in turn, abused by unreasonable minority shareholders.[61] In other words, fair remedies will not threaten normal corporate governance when applied.

III. The Heart of the Unfair Prejudice Remedy: Broad Judicial Discretion

The statutory framework for the unfair prejudice petition in the UK is set out in s 994 and s 996 of the CA 2006. The courts are given wide-ranging discretion to determine what conduct unfairly prejudices the petitioner under s 994[62] and what relief should be granted to the petitioner under s 996.[63] In this sense, it has been suggested that the effectiveness of the unfair prejudice remedy regime depends on the creativity of judges in interpreting and applying s 994 and s 996.[64] Thus, Section 3 of this article will review how the court's interpretive discretion has worked from the perspective of s 994 and s 996 respectively.

58 DK Moll, 'Shareholder Oppression and the New Louisiana Business Corporation Act' (2015) 60 *Loyola Law Review* 461, 465.

59 *Thomas v H W Thomas Ltd* (1984) 1 NZLR 686.

60 ibid at 694–695.

61 *Re Ring Tower (No 2)* [1989] BCLC 427.

62 Companies Act 2006, s 994.

63 Companies Act 2006, s 996.

64 J Lowry, 'The pursuit of effective minority shareholder protection: s.459 of the Companies Act 1985' (1996) 17 Company Lawyer 67, 67.

A. SECTION 994 of CA 2006: BROAD SCOPE OF APPLICATION

Section 994 of the CA 2006 provides that the grounds for a petition are that "the company's affairs are being or have been conducted in a manner that is unfairly prejudicial to the interests of members generally or of some part of its members".[65] The scope of application of this provision primarily relates to how the courts interpret the interrelated concepts of 'unfairly prejudicial' and 'interest'.[66] The wording itself seems sufficiently open to leave room for extensive interpretation, but the case law can reflect efforts to balance the discretion of the courts with legal certainty.

(i) Term: Unfairly Prejudicial

To ensure the flexibility of s 994, the CA 2006 does not make a comprehensive definition of 'unfairly prejudicial', but its guiding principles have been developed in the case law.[67] Nevertheless, before discussing how the term has been interpreted, it is necessary to note that the court needs to rely on the standard of objectivity in determining whether the act complained of has been unfairly prejudicial.[68] This standard does not require the petitioner to prove that the respondent had a malicious intent to cause harm.[69] Instead, the court tends to focus on the actual impact of the misconduct on the petitioner.[70] On this basis, the complained conduct must satisfy both the 'unfairness test' and the 'prejudice test'.

In the first place, the 'unfairness test', centres on assessing whether the principles of equality and good faith can be superimposed on the exercise of legal rights.[71] In most situations, the corporate structure consisting of the Companies Act and the articles of association is adequate and exhaustive,[72] and the latter in particular are deemed to be the result of prior bargaining between the parties over the efficient use of resources.[73] Given the importance of commercial practice emphasising compliance with commitments and agreements, equitable considerations cannot often easily override the articles of association or subsidiary agreements between members of a company.[74] As a consequence, a judge typically

[65] Companies Act 2006, s 994(1).
[66] Law Commission, (n 1) para 9.17.
[67] Z Fan, 'Unfair prejudice in United Kingdom Company Law' (2021) 9 Asian Journal of Humanities and Social Studies 27, 28.
[68] N Flourentzou, (n 8).
[69] *Re Saul D Harrison & Sons plc* [1995] 1 BCLC 14, at 31.
[70] DD Prentice, 'The Theory of the Firm: Minority Shareholder Oppression: Sections 459-461 of the Companies Act 1985' (1988) 8 Oxford Journal of Legal Studies 55, 78.
[71] JJ Chapman, (n 26) 179.
[72] *Ebrahimi v Westbourne Galleries Ltd* [1972] 2 All ER 492, at 500.
[73] BR Cheffins, *Company Law: Theory, Structure and Operation* (Oxford University Press 1997), 274–275.
[74] *Re Saul D Harrison,* (n 69) at 18.

shapes the concept of 'fairness' from a judicial perspective based on the 'reasonable principle' rather than making any order that he considers fair based on his own value judgement.[75]

Notably, *O'Neill v Phillips*, the only case on the unfair prejudice clause currently before the House of Lords, involved a convincing explanation of the 'reasonable principle'.[76] Specifically, Lord Hoffmann has developed a framework for the 'unfairness test': the first step is to determine whether the applicant has breached agreed terms for the conduct of the company's affairs (for example, the articles of association or any collateral agreements between shareholders); if not, the second step is to judge whether the respondent has acted in a manner contrary to the principles of good faith relevant to equity (*i.e.*, equitable considerations may in some circumstances render unfair the exercise of strict legal powers under the company's constitution).[77] Fundamentally, without eliminating contractual arrangements between members, the 'unfairness test' provides opportunities to moderate or limit the exercise of contractual rights when enforcement of those rights would be unconscionable.[78] Accordingly, members are likely to petition against a strict infringement of a legal right or an unfair use of power.[79] This indicates that the concept of 'fairness' under s 994 cuts across the distinction between acts of legality and illegality.[80]

Nonetheless, an issue that must be mentioned is whether the 'clean hands' rule affects the 'unfairness test', *i.e.*, whether the petitioner's own misconduct would prevent the application of established equitable principles.[81] As noted earlier, unfair prejudice is an objective matter of judgment, so in theory the petitioner is not required to come to court with clean hands. However, a court probably denies relief to a petitioner if his conduct was grossly improper[82] or if his conduct was closely related to the respondent's unfair prejudice.[83] Therefore, the 'clean hands' rule is a consideration for the court when examining the concept of 'fairness', which may, to some extent, limit the potential abuse of the term.

In the second place, as reliance on the 'unfairness test' alone may lead to over-protection of minority shareholders by the law, the 'prejudice test', which requires the petitioner to suffer some form of loss before relief can be obtained,

75 *O'Neill v Phillips* [1999] 1 WLR 1092, at 1098.
76 ibid.
77 ibid.
78 *Richard Moxon v Litchfield, Cook, Kulesza* [2013] EWHC 3957.
79 FF Ma, 'A Comparative Analysis of Minority Shareholders' Remedies in Anglo-American Law and Chinese Law: Lessons to be Learnt' (*PhD, University of the West of England* 2009), 178.
80 *Re A Company (No. 8699 of 1985)* [1986] BCLC 382 at 387.
81 B Hannigan, (n 20) ch 19, 508.
82 *Interactive Technology Company Limited v Jonathan Ferster and Ors* [2016] EWHC 2896, at 318–325.
83 *Re A Noble & Sons (Clothing) Ltd* [1983] BCLC 273.

provides a reasonable basis for judicial intervention.[84] Nevertheless, the notion of 'prejudice' is equally broad, with its division into financial and non-financial prejudice. On the one hand, economic prejudice usually means that the value of a member's shares in a company has been seriously jeopardised by the actions of those who have substantial control over the company.[85] Financial prejudice is also likely to include other financial losses connected with the petitioner's status as a member.[86] For instance, where a member has an equitably recognised right to the management of the company, the exclusion of that member from the corporate management and the resulting loss of income or profits from the company in the form of remuneration would constitute prejudice.[87] On the other hand, if a member's rights are disregarded, the 'prejudice test' may be triggered, even if there are no financial consequences.[88] Taking *Quinlan v Essex Hinge Co Ltd*[89] as an example, the minority shareholder Mr. Quinlan was dismissed as a director by the controlling shareholder Mr. Reid. Then, Mr. Quinlan repeatedly asked Mr. Reid about the reasons for his removal but received no response, which could be understood as non-financial prejudice from Mr. Reid.[90]

(ii) Term: Interest

The 'interest' that s 994 seeks to protect is not every interest of the petitioner, but his interest as a member of the company.[91] Nonetheless, the legislator's use of the word 'interest' rather than 'right' creates scope for members to accommodate a wider range of complaints than those based on strict legal rights.[92] Consequently, simply asking about the identity of the actor is not sufficient to clarify the meaning of 'interest', which needs to be considered in conjunction with the notion of 'fairness'.[93] In an equitable position, individual members' "rights, expectations and obligations inter se which are not necessarily submerged in the company structure[94]". In this way, 'interests' include not only the legal rights

84 PMC Koh, 'A Reconsideration of the Shareholder's Remedy for Oppression in Singapore' (2013) 42 *Common Law World Review* 61, 74.

85 *Re A Noble & Sons,* (n 83) at 290-291.

86 V Joffe QC, D Drake, G Richardson, D Lightman QC, and T Collingwood, 'Unfair Prejudice: The Statutory Remedy' in *Minority Shareholders: Law, Practice, and Procedure* (6th edn, Oxford University Press 2018), ch 6, 330.

87 *McKillen v Misland (Cyprus) Investments Ltd* [2012] EWHC 2343 (Ch), [630], [2013] 2 BCLC 583.

88 ibid.

89 *Quinlan v Essex Hinge Co Ltd* [1996] 2 BCLC 417.

90 ibid.

91 Law Commission, (n 1) para 9.20.

92 AO Nwafor, 'Unfair Prejudice Remedy: A Relief for the Minority Shareholders - A Comparative Perspective' (2011) 22 *International Company and Commercial Law Review* 285, 289–290.

93 ibid.

94 *Ebrahimi v Westbourne Galleries Ltd* [1973] AC 360, at 379.

of minority shareholders by virtue of the company's constitution or shareholders' agreement, but also their 'legitimate expectations'.[95]

Legitimate expectations commonly "arise out of a fundamental understanding between the shareholders which formed the basis of their association but was not put into contractual form".[96] That is to say, in determining legitimate expectations, the court should concentrate on the relationship between the shareholders and the existence of informal agreements or arrangements outside the constitution.[97] Significantly, legitimate expectations are considered more likely to exist in the 'quasi-partnership', namely family-owned businesses with strong private attributes.[98] In this regard, Lord Wilberforce in *Ebrahimi v Westbourne Galleries Ltd* enumerated three essential features of a 'quasi-partnership' - (1) a personal relationship of mutual trust as the basis of a business association; (2) an agreement, commitment or understanding that members will participate in the management; and (3) a restriction on the transfer of shares to prevent members leaving.[99] These elements correspond to a large extent to the private companies mentioned earlier in Section 2.1. Most private companies are formed in an atmosphere of partnership and trust, so that shareholders develop a reasonable reliance on obtaining a return on their investment and participating in the management of the company, even though these matters may not be spelled out in the articles or other subsidiary agreements. In this sense, s 994 petitions are welcomed by members of such companies, as legitimate expectations probably cover anything beyond the strict language of the contract.[100]

Nevertheless, Lord Hoffmann in *O'Neill* refused to rely on 'legitimate expectations' as a basis for a claim because s 994 does not offer the court the general power to assess the fairness of the conduct by majority shareholders.[101] Considering the risk that the liberal position represented by 'legitimate expectations' would open the floodgates for s 994 petitions,[102] he preferred to use 'equitable considerations' to describe the foundation for judicial intervention against unfairly prejudicial conduct.[103] Technically, building on the *Ebrahimi* rule, the *O'Neill* decision stresses the importance of the traditional equitable principles and contractual doctrine to assess whether there has been some infringement of

95 Law Commission, (n 1) paras 9.17–9.20.
96 *Re Saul D Harrison*, (n 69) at 18–19.
97 DD Prentice, (n 70) 59.
98 *Re A company (No 00477 of 1986)* [1986] BCLC 376, at 378.
99 *Ebrahimi*, (n 72).
100 JJ Chapman, (n 26) 207.
101 D Ohrenstein, 'Minority Shareholders & Unfair Prejudice' (*Radcliffe Chambers*, 2011) <https://radcliffechambers.com/wp-content/uploads/2019/11/Minority_Shareholders_and_Unfair_Prejudice_Lecture-DO.pdf> accessed 28 July 2021.
102 B Clark, 'Unfairly Prejudicial Conduct' (1999) 38 Scots Law Times 321, 323.
103 *O'Neill v Phillips* [1999] 2 BCLC 1, at 11.

the petitioner's formal or informal rights.[104] Arguably, this restriction of the concept of 'legitimate expectations' expresses concern about the overly broad discretion of the courts and reflects the move towards greater certainty in s 994 petitions.[105]

B. Section 996 of CA 2006: Diversified Remedies

If the court is satisfied that the unfair prejudice petition presented is well-founded, it may make such order under s 996 of the CA 2006 as it thinks fit to provide relief to the petitioner.[106] In exercising that discretion, the court should consider all relevant factors that may affect the relief.[107] Basically, a remedy must be proportionate to the unfair prejudice found.[108] If the consequences of the unfairly prejudicial conduct are not severe, it is relatively inappropriate to enforce some potentially drastic remedies.[109] Moreover, in considering the interests of litigants, the court cannot turn a blind eye to the interests of stakeholders or the company itself, although the weight to be given to their interests will depend on the circumstances.[110] In VB *Football Assets v Blackpool Football Club (Properties) Ltd*, for instance, the interests of the football club were deemed to be a crucial consideration in the court's decision as to what type of order to make under s 996.[111] In *Re Asia Television Ltd*, in the context of a provision equivalent to s 994, Harris J held that where the nature of the company's activities is public in nature, it is necessary to take into account the interests of the company as a whole, its creditors, its employees and the public in granting relief.[112]

S 996(2) provides a detailed list of the types of remedies available to the court, including buy-out orders, regulation of the affairs of the company, and injunctive relief, etc.[113] Essentially, buy-out orders[114] and authorisation for shareholders to bring derivative actions[115] are two common ways in which courts and shareholders are concerned. In the case of buy-out orders (*i.e.,* requiring the company or respondent to purchase the petitioner's shares), the Law Commission

104 C Newington-Bridges, 'A Practical Guide to Unfair Prejudice Petitions and their interaction with Derivative Claims' (*St John's Chambers*, 2016) <https://www.stjohnschambers.co.uk/wp-content/uploads/2018/07/Unfair-prejudice-petitions-and-derivative-actions.pdf> accessed 10 July 2021.

105 MA Iqbal, (n 5) 184.

106 Companies Act 2006, s 996(1).

107 *Grace v Biagioli* [2006] 2 BCLC 70, at 73.

108 V Joffe QC, D Drake, G Richardson, D Lightman QC, and T Collingwood, (n 86) ch 7, 422.

109 *Re Phoenix Office Supplies Ltd* [2003] 1 BCLC 76, at 51.

110 V Joffe QC, D Drake, G Richardson, D Lightman QC, and T. Collingwood, (n 86) ch 7, 425.

111 *VB Football Assets v Blackpool Football Club & Ors* [2017] EWHC 2767 (Ch) at 447.

112 *Re ATV Television Ltd* [2015] 1 HKLRD 607, at 55–56.

113 Companies Act 2006, s 996(2).

114 ibid s 996(2)(e).

115 ibid s 996(2)(c).

found it to be the most attractive remedy after a statistical survey of unfair prejudice cases.[116] The advantage of this approach is that it provides a judicially created exit for the shareholder, allowing him to recover the capital he has invested in the business without dissolving the company.[117] From this perspective, a buy-out order is a desirable measure to safeguard the interests of the petitioner, the respondent and the company.

However, the court's power to authorise the petitioner to bring a separate derivative action under s 996(2)(c) is a controversial topic. Shareholders have been subject to the proper plaintiff rule highlighted in *Foss v Harbottle* and therefore cannot allege in their own name that a member has committed a wrong against the company[118], unless they satisfy the requirements of a statutory derivative action under Part 11 of the CA 2006.[119] In theory, the effect of s 996(2)(c) is to enable the petitioner to overcome some of the obstacles inherent in bringing a derivative action.[120] Nevertheless, the petitioner must first incur additional costs and time to prove the existence of unfair prejudice before obtaining the court's authorization.[121] In such circumstances, it is difficult to see why two sets of procedures would be more cost-effective than a court granting relief directly to the company.[122] In addition, unlike most other orders that might be made under s 996, a derivative action authorised under s 996(2)(c) would ultimately benefit the company, not the individual shareholder.[123] Hence, the application of s 996(2)(c) may be limited in practice.

IV. THE EXISTING DILEMMA OF THE UNFAIR PREJUDICE REMEDY: CONFUSED BY THE UNCERTAINTY OF JUDICIAL DISCRETION?

The flexibility of unfair prejudice remedies is evidenced by the expansive interpretation of certain terms in s 994 and the wide range of remedies provided by s 996, which are considered to provide adequate protection to minority shareholders.[124] Also, given the potential for cunning and opportunistic use of s 994 by minority shareholders, the court was mindful of the need to adopt a more measured response to interference in the affairs of the company.[125] Nonetheless,

116 Law Commission, (n 14) para 3.3.
117 *Grace*, (n 107) at 75.
118 *Foss v Harbottle* (1843) 2 Hare 461.
119 Companies Act 2006, part 11.
120 V Joffe QC, D Drake, G Richardson, D Lightman QC, and T Collingwood, (n 86) ch 7, 429.
121 Law Commission, (n 1) paras 10.8-10.9.
122 J Payne, 'Sections 459-461 Companies Act 1985 in Flux: The Future of Shareholder Protection' (2005) 64 Cambridge Law Journal 647, 653.
123 V Joffe QC, D Drake, G Richardson, D Lightman QC, and T Collingwood, (n 86) ch 7, 430.
124 N Flourentzou, (n 8).
125 J Mukwiri, (n 11) 283–284.

the inherent vagueness of the language of s 994 leaves the courts with uncertainty in the exercise of their judicial discretion. This has led not only to lengthy procedures for s 994 petitions, but also to an overlap between s 994 and Pt 11 jurisdiction of the CA 2006. Section 4 of this article will critically analyse how these two adverse consequences prevent the unfair prejudice remedy system from achieving the goals of 'efficiency' and 'fairness' referred to in the previous Section 2.3.

A. The burdensome proceedings for s 994 petitions: a time-consuming and costly process

Due to the broad scope of s 994, a petitioner may raise any fact relevant to the management of a company's business.[126] This is likely to cause "complex, often historical, factual investigations" and "costly, cumbersome litigation".[127] In short, the time and cost challenges of unfair prejudice proceedings can place additional burdens on courts, litigating shareholders, stakeholders and companies.

Firstly, as Hoffmann J noted in *Re Unisoft Group Ltd (No 3)*, unfair prejudice petitions are 'notorious', particularly for the courts and potential parties to such proceedings, because of the length and unpredictability of the management of these cases, which often incur appalling judicial costs.[128] In practice, if the petitioner proves that the *Ebrahimi* test is satisfied, the likelihood of success will increase.[129] This seems to encourage the parties to provide a detailed account of the history of the company and the understandings and agreements reached between them.[130] However, examining matters that may have occurred many years ago can be problematic for the court, especially in the case of *Re Macro (Ipswich) Ltd* which involved a historical investigation into the affairs of the company spanning approximately 40 years.[131] Furthermore, the large amount of relevant evidence probably has contributed to the vagueness and imprecision of the petition, and some of the matters alleged were not even within the scope of s 994, which largely prolonged the court's consideration of the case.[132] Likewise, litigating shareholders are probably subject to financial pressure. For example, in *Re Elgindata Ltd*, the hearing lasted 43 days, cost £320,000, and the petitioner's shares were valued at £24,600, down from £40,000 at the time of purchase.[133]

[126] *Re Sam Weller & Sons Ltd* [1990] Ch 682 (Ch).
[127] Law Commission, (n 1) para 14.5.
[128] *Re Unisoft Group Limited (No 3)* [1994] 1 BCLC 609, at 611.
[129] B Hannigan, (n 20) ch 19, 517–519.
[130] Law Commission, (n 1) para 11.10.
[131] *Re Macro*, (n 43).
[132] *Re Unisoft Group,* (n 128).
[133] ibid.

Secondly, it is important to mention that companies and stakeholders are possibly caught up in the onerous proceedings of s 994 petitions. In the case of companies, for one thing, prolonged hearings are likely to distract the company's management and thus adversely affect the company's day-to-day operations.[134] For another, the company's reputation may be corroded as its assets and operations may be frozen or severely restricted during the proceedings.[135] Additionally, the financial hardship caused to the company by a s 994 petition could expose stakeholders to potential losses. For instance, other shareholders who do not file a petition may find that their profits and stock prices fall, or that creditors may find it hard to collect amounts normally due from the company.

B. Overlapping jurisdiction between section 994 and statutory derivative claims: corporate wrong

A subject of concentrated academic debate at present is whether an action for unfair prejudice under s 994 should be used to deal with corporate wrongs.[136] Traditionally, corporate claims have fallen within the scope of statutory derivative jurisdiction,[137] whereas s 994 is a personal remedy for individual shareholder claims.[138] Nevertheless, recent case law has revealed a trend where alleged breaches of directors' duties can establish a claim under s 994.[139] This indicates that s 994 petitions may not be limited to claims of a personal nature and thus their jurisdictional scope may have been further expanded. While this expansive approach to interpretation provides greater convenience to minority shareholders from an efficiency perspective, it probably increases the risk that they will abuse s 994 to pursue vexatious claims against the company.[140]

It is necessary to clarify that the remedy of authorising derivative claims under s 996(2)(c), as described in Section III.B above, is different from bringing a corporate claim under s 994, as discussed here. The former is the result of a successful unfair prejudice action, whereas the latter emphasises that corporate wrong are construed as the cause of such action.

[134] Z Fan, (n 67) 36.
[135] Law Commission, (n 1) para 11.2.
[136] PI Roberts and J Poole, (n 10).
[137] A Gray, 'The Statutory Derivative Claim: An outmoded superfluousness?' (2012) 33 Company Lawyer 296, 298.
[138] J Payne, 'Shareholders' Remedies Reassessed' (2004) 67 Modern Law Review 500, 501–503.
[139] For example, *Re Cumuana Ltd* [1986] BCLC 430; *Re McCarthy Surfacing Ltd* [2009] 1 BCLC 622.
[140] FF Ma, (n 79) 184.

(i) Personal and Corporate Relief: A Progressively Blurred Boundary

In the context of English company law, directors owe a fiduciary duty to the company, so a breach of a director's duties is regarded as a wrong committed against the company rather than the shareholder.[141] If a shareholder wishes to sue a director for wrongdoing, it will need to commence derivative proceedings under Pt 11 of the CA 2006 to exercise the company's rights.[142] In contrast, the core of s 994 lies in the personal rights of shareholders.[143] As highlighted by Millet J in *Re Charnley Davies Ltd (No 2)*, a clear distinction needs to be maintained between corporate remedies, which can be obtained through derivative proceedings, and individual remedies, which can be obtained through unfair prejudice proceedings.[144]

Recently, however, courts have tended to adopt a more liberal interpretation in favour of using s 994 to seek corporate relief. For example, Lord Hoffmann in *Re A Company (No. 005287 of 1985)* considered the situation where a successful unfair prejudice petition denied corporate relief and authorised the plaintiff to commence a derivative action at that stage.[145] He claimed that this could lead to unnecessary duplication of litigation.[146] In particular, in the landmark case of *Clark v Cutland*, the Court of Appeal confirmed that minority shareholders are permitted to use the unfair prejudice clause to obtain substantive relief for corporate wrongs.[147] It is fair to say that this decision blurs the traditional boundary between corporate wrongs remedied by derivative actions and personal wrongs remedied by s 994 actions.[148]

Indeed, the *Cutland* decision is to some extent logical and cost-effective. Firstly, the language of s 994 does not limit its application exclusively to unfair prejudice in the form of infringement of the individual rights of shareholders.[149] As noted earlier in Section 3.1.2, 'interest' under s 994 is a broad and flexible term. The term makes it clear that any wrongful conduct prejudicial to the interests of a member of a company, including a breach of a director's duties or other wrongful conduct towards the company, will be governed by s 994.[150] Similarly, there is no a priori reason to exclude company-related relief from unfair prejudice petitions.

141 *Percival v Wright* [1902] 2 Ch 421.
142 Companies Act 2006, (n 120).
143 *Re Saul D Harrison*, (n 69).
144 *Re Charnley Davies (No 2)* [1990] BCLC 760, at 784.
145 *Re A Company (No 005287 of 1985)* [1986] 1 WLR 281.
146 ibid, at 284.
147 *Clark v Cutland* [2003] 4 All ER 733.
148 C Hirt, 'In what Circumstances should Breach of Director's Duties Give Rise to A Remedy under ss 459–461 of the Companies Act 1985?' (2003) 24 Company Lawyer 100, 106-109.
149 DD Prentice, n (70) 66.
150 R Cheung, 'Corporate Wrongs Litigated in the Context of Unfair Prejudice Claims: Reforming the Unfair Prejudice Remedy for the Redress of Corporate Wrongs' (2008) 29 *Company Lawyer* 98, 101–103.

For instance, although the payment of excessive director's remuneration[151] and the improper transfer of the company's business[152] both involve a breach of a director's duty to the company, case law recognises these acts as unfairly prejudicial. Secondly, since the court can authorise derivative actions under s 996(2)(c), corporate relief is in principle the appropriate outcome of an unfair prejudice petition, even though such a remedy is procedural rather than substantive in form.[153] In this sense, it is demanding to explain why allegations of corporate wrong should be permitted while denying appropriate relief to the corporation.[154] Besides, applying s 994 to corporate wrongs would allow individual shareholders to bypass the procedural requirements of derivative claims and therefore save their time and costs. In summary, there appears to be good reason to support an unfair prejudice action to redress the wrongs done to the company.

(ii) Minority Shareholders versus Companies and Majority Shareholders: A Tilted Balance of Interests

While the flexible *Cutland* approach has widened the scope of shareholder relief, it still leaves substantial uncertainty and ambiguity.[155] In exercising the judicial discretion regarding s 994 petitions, the importance of preserving the proper balance between the interests of minority shareholders and corporate autonomy needs to be borne in mind. Also, an acceptable remedy for unfair prejudice cannot be at the expense of the reasonable interests of the majority shareholder. Nonetheless, the overlap of jurisdiction between s 994 and derivative claims may tip the judicial scales in favour of minority shareholders. This raises alarm as to whether minority shareholders may abuse the sympathies of the court and cause unnecessary problems for the company and the majority shareholders.

In the first place, the procedural limitations of s 994 itself do not offer sufficient protection for companies as compared to derivative actions.[156] In essence, the unfair prejudice remedy regime focuses on resolving disputes between shareholders and does not provide a basis for determining whether it is in the best interests of the company to pursue a claim on its behalf under s 994.[157] Rather, derivative proceedings designed to do justice to companies, as evidenced by its well-designed leave threshold[158] and two-stage procedural threshold[159] to

151 *Re Cumana Ltd* [1986] BCLC 430.
152 *Re Stewarts (Brixton) Ltd* [1985] BCLC 4.
153 R Cheung, (n 150).
154 PI Roberts and J Poole, (n 10) 120.
155 FF Ma, (n 79) 189.
156 ibid, 190.
157 J Payne, (n 122) 660.
158 Companies Act 2006, s 261(1).
159 ibid s 263(2)-(4).

avoid opening the floodgates. Thus, in contrast to derivative actions, s 994 lacks a sophisticated mechanism for conducting or controlling litigation relating to corporate remedies. As an example of the treatment of malicious cases, in derivative claims, the court would exercise strike-out jurisdiction at the leave stage. However, in s 994 petitions, it is unclear whether the court would strike out frivolous claims at an early stage or a full trial, depending on whether the defendant files a motion to strike out.[160] Moreover, s 994 jurisdiction does not require the court to consider factors such as whether the misconduct has been approved[161] or whether an independent body within the company wishes to bar the action.[162] Obviously, the s 994 means of screening out improper conduct is essentially inadequate to protect companies from malicious interference by petitioners.

In the second place, when the petitioner seeks to seek personal relief for corporate wrongs under s 994, the respondent may be at risk of double recovery.[163] In fact, this argument relates to the applicability of the 'no reflective loss' principle to s 994 claims.[164] The principle is based on derivative claims, which preclude shareholders from bringing a personal claim for a reduction in the value of their shareholding as a result of a director's wrongful conduct towards the company.[165] This is because the affected shareholders can recover their losses when the company exercises its right to relief under the *Foss* rule.[166] Hence, the application of the 'no reflective loss' principle helps to avoid shareholders receiving double compensation for wrongdoing directors. Nevertheless, in the context of unfair prejudice, *Atlasview Ltd v Brightview Ltd*[167] set a precedent for the admissibility of personal remedies for corporate wrongs. The High Court in *Atlasview*, after reviewing past jurisprudence, held that there was no valid basis for stating that the 'no reflective loss' argument created a barrier to the relief sought in an unfair prejudice petition.[168] Nonetheless, it is debatable to what extent the fact that the argument was 'not raised in the past' is a compelling reason for a court to refuse to apply the argument to a s 994 action.[169] Furthermore, the *Atlasview* judge relied heavily on pre-*Johnson v Gore Wood* precedents[170] which did not address the issue

160 J Payne, (n 122) 659.
161 Companies Act 2006, s263(2)(c).
162 ibid, s 263(4).
163 S Perera, 'Reconceptualising Shareholder Remedies to Mitigate the Problems Caused by the Overlap between Section 994 and Part 11 Companies Act 2006' (2019) 8 *UCL Journal of Law and Jurisprudence* 1, 9.
164 FF Ma, (n 79) 189.
165 Prudential, (n 19) at 366–367.
166 ibid.
167 *Atlasview Ltd v. Brightview Ltd* [2004] 2 BCLC 191, at 208.
168 ibid.
169 J Payne, (n 122) 669.
170 For example, *Re A Company (No 003843 of 1986)* [1986] BCLC 68; Saul D Harrison, (n 69).

of reflective loss.[171] Notably, the *Johnson* decision, in which the circumstances of reflective loss were fully considered by the House of Lords, was not paid attention.[172] Accordingly, the line of reasoning of the *Atlasview* court may be less than appropriate. At least, If accepted as authority, the *Atlasview* decision is a troubling sign for the controlling directors, suggesting that a s 994 petition could be a shortcut to double damages for shareholders. In this way, unfair prejudice claims may become a tool for the minority to oppress the majority.

V. REFORM OF THE UNFAIR PREJUDICE REMEDY: GUARANTEEING THE APPROPRIATENESS OF JUDICIAL DISCRETION

Given the complex interpersonal relationships that exist in private companies, judicial intervention is inevitable and its uncertainty is a fair price to pay for providing flexible remedies to combat the opportunistic behaviour of shareholders.[173] Nonetheless, too much ambiguity would allow the scope of unfair prejudice remedies to be extended so far that it could subsume the whole corporate law.[174] Therefore, the challenge for s 994 petitions is to keep the court's discretion within reasonable limits to balance cost-effectiveness and equity considerations. Based on two guiding criteria mentioned in Section 2 above and the dilemma of s 994 petitions mentioned in Section 4 above, Section 5 of this article aims to aims to propose a framework for judicial discretion in two respects: firstly, by creating a statutory presumption method in relation to the determination of unfair prejudicial conduct, which is conducive to reducing the time and cost of s 994 proceedings; and secondly, by making a conditional distinction between s 994 action and derivative action jurisdiction, which is conducive to striking a balance between the interests of shareholders and those of the company.

A. APPROACHES TO PROMOTING EFFICIENCY: TWO STATUTORY PRESUMPTIONS

To help the courts address the length, complexity and cost problems of unfair prejudice actions, the English Law Commission recommended the adoption of two rebuttable statutory presumptions - treating the exclusion of shareholders from management as unfairly prejudicial and granting specific relief where certain

[171] J Payne, (n 122) 669.
[172] ibid.
[173] S Miller, 'How should UK and US Minority Shareholder Remedies for Unfairly Prejudicial or Oppressive Conduct be Reformed?' (1999) 36 American Business Law Journal 579, 613.
[174] JJ Chapman, (n 26) 171.

conditions will be met.[175] However, these recommendations were rejected by the CLRSG and are thus not reflected in the CA 2006. Against this background, this article challenges the CLRSG's opposing position and argues that the presumption approach is largely a reasonable measure to inject appropriate certainty into an excessively flexible judicial discretion.

The statutory presumption method is primarily aimed at private companies in which all or almost all members are directors.[176] Particularly, members who hold at least 10% of the voting rights in their own name will be eligible to petition.[177] Since the removal of a shareholder from management and a buy-out order are the most common causes of action for unfair prejudice and the most commonly sought remedy respectively,[178] the proposed approach falls into two main presumptions: firstly, where a shareholder is excluded from the management of a company, for example, if he is removed as a director or otherwise prevented from performing the functions of a director, the act will be presumed to unfair prejudice unless there is evidence to the contrary.[179] Secondly, if the first presumption is not rebutted and the court is satisfied that it is necessary to order the respondent to buy out the petitioner's shares, the shares should be valued pro rata unless the court orders otherwise.[180]

As discussed in Section IV.A, courts frequently have to consider a large number of factual allegations in unfair prejudice petitions. In such cases, the Law Commission stated that using the statutory presumptions would provide greater certainty to the parties at the time of litigation, thereby allowing the case to be dealt with more quickly.[181] For instance, where circumstances arise in relation to the first presumption, the defendant may rebut it through introducing evidence to which the presumption should not apply, thus limiting the scope of the court's historical inquiry.[182] However, the CLRSG was concerned that the reform measure probably encourages litigation.[183] Under the proposed conditions, as it may be imprudent to directly presume that a decision to remove a shareholder from management is unfair or to treat a buyout order directly as an appropriate remedy, the CLRSG questioned the potential for abuse of the proposed statutory presumptions.[184]

175 Law Commission, (n 14) para 3.30.
176 ibid, para 3.48.
177 ibid, paras 3.45–3.47.
178 ibid, para 3.3.
179 ibid, para 3.56.
180 ibid, para 3.62.
181 ibid, para 3.28.
182 ibid.
183 Department of Trade and Industry, Developing the Framework (2000) *URN 00/656*, para 4.104 <https://webarchive.nationalarchives.gov.uk/ukgwa/+/http://www.berr.gov.uk/whatwedo/businesslaw/co-act-2006/clr-review/page25086.html> accessed 20 September 2021.
184 ibid.

Unfortunately, the CLRSG is likely to exaggerate the shortcomings of the statutory presumption approach. Firstly, the CLRSG may overlook the fact that the greatest strength of the approach is that it is founded on 'structural' factors rather than on vague 'expectations or understandings' between the parties.[185] These 'structural' matters, such as the petitioner's shareholding and the fact that he is a director, can be readily determined by reference to the recent position.[186] Hence, there is an opportunity for the court to be freed from the cumbersome fact-finding process, which probably facilitates the efficiency of the hearing of a s 994 petition. Also, lawyers representing the parties can tell their clients with greater certainty about their prospects of success.[187] As a consequence, Boyle correctly argues that the presumption approach probably provides a more predictable process to increase the cost-effectiveness of the court and the parties.[188] Significantly, the statutory presumptions would prompt more unfair prejudice claims to be settled out of court or before the hearing, without opening the floodgates to such claims.[189] Secondly, the proposed presumptions do not lose flexibility by adding certainty to s 994 petitions.[190] While the presumptions built on 'structural' factors may seem somewhat arbitrary, if a case does not meet the conditions under which the presumptions arise, the application of s 994 is not affected by the absence of the presumptions.[191] Additionally, even if the presumption applied, the court might still find that it was not unfair to exclude the petitioner from management, or allow the respondent to purchase the petitioner's shares at a discount, if the respondent adduced evidence to the contrary.[192] In other words, the statutory presumptions merely provide a potential way to alleviate the difficulties of lengthy and costly s 994 litigation, but the court still has full discretion to determine the existence of unfairly prejudicial conduct and to decide what remedy should be granted. Overall, the benefits of the statutory presumptions for the efficiency of unfair prejudice actions probably outweigh its limitations.

185 Law Commission, (n 14) para 3.37.
186 ibid.
187 R Cheung, 'The Statutory Minority Remedies of Unfair Prejudice and Just and Equitable Winding up: the English Law Commission's Recommendations as Models for Reform in Hong Kong' (2008) 19 International Company and Commercial Law Review 156, 162–163.
188 AJ Boyle, Minority Shareholders' Remedies (Cambridge University Press 2002), 125–126.
189 ibid.
190 FF Ma, (n 79) 205.
191 Law Commission, (n 14) para 3.37.
192 ibid.

B. Approaches to Promoting Fairness: A Reasonable Distinction Between the Jurisdiction of s 994 and Derivative Actions

The fundamental contradiction in the overlap of jurisdiction between s 994 of the CA 2006 and derivative actions is that the former does not have the appropriate procedural thresholds and 'no reflection loss' principles to screen out frivolous and worthless shareholder actions that the latter does. If s 994 were to be broadly extended to cover corporate remedies, fair legal mechanisms would need to be put in place to ensure that s 994 is not abused by minority shareholders.[193] Nevertheless, the Law Commission has not fully addressed this issue in its review of shareholder remedies.[194] Therefore, Section 5.2 of this article will attempt to present a legal framework that applies to the overlapping dilemma.

(i) Scenario 1: Using s 994 to Obtain Corporate Relief on A Corporate Claim

The opening statement of Section 5.2.1 is that the *Cutland* approach, which allowed the court to automatically order corporate relief after determining the criteria for a petition, should be abandoned.[195] Instead, the court should have absolute discretion to deny substantive relief to the company where it is appropriate to do so.[196] Consequently, it is necessary to place some procedural hurdles or considerations in a s 994 petition so that the court can decide in advance whether to enable shareholders to bring claims on behalf of the company through s 994.

Firstly, the two-step framework summarised by Perera in light of *Charnley Davies*[197] and *Chime Corp*[198] decisions is informative. Lord Millett's comments in *Charnley Davies* is the starting point for the court's jurisdiction to distinguish between unfair prejudice claims and derivative claims.[199] The first step of the framework requires the court to examine all the elements of the claim at the pleading stage.[200] Specifically, the court is tasked with determining whether the content of the unfair prejudice petition is essentially 'misconduct' or 'mismanagement'.[201] Lord Millett explains that the difference between the two lies

[193] R Cheung, (n 150) 103–104.
[194] PI Roberts and J Poole, (n 10) 44.
[195] S Perera, (n 163) 21.
[196] R Cheung, (n 150) 104.
[197] *Charnley,* (n 144) at 783.
[198] *Chime Corp, Re* [2004] 7 HKCFAR 546, at 62–63.
[199] S Perera, (n 163) 21.
[200] *Charnley,* (n 144), at 783.
[201] ibid.

in the 'nature of the complaint' and the 'appropriate remedy'.[202] In terms of the 'nature of the complaint', 'misconduct' refers only to unlawful conduct, such as breach of directors' duties, whereas 'mismanagement' relates to a broader category of unlawful conduct (*i.e.,* potentially including misconduct 'in part but not in whole').[203] Arguably, if the court is satisfied that the petition presented satisfies the elements of 'misconduct', then the derivative claim must apply.[204] On the contrary, if the petitioner has suffered loss as a result of the directors' mismanagement, then in principle personal, rather than corporate, relief can be sought under an unfair prejudice petition.[205] Although Lord Millett recognises that two separate claims can be created on the same facts, the 'nature of the complaint' and the 'appropriate remedy' are different in the two cases.[206]

Next, the second step can be found in Lord Scott's statement in the Hong Kong case of *Chime Corp*.[207] The legal background of that case is comparable to that of s 994. Basically, Lord Scott endorsed Lord Millett's distinction between 'misconduct' claims and 'mismanagement' claims.[208] More critically, he further adds to the framework of this categorisation by arguing that the court may exercise discretion in dealing with 'mismanagement' claims to allow the petitioner to obtain corporate relief subject to overcoming these two hurdles: firstly, the need to establish the value of directors' liability at the pleading stage.[209] Secondly, the relief ordered needs to be consistent with the remedy available if a derivative claim is established.[210]

While the legal framework consisting of the above two steps has not been formally applied, it can be justified in some cases where relief has been ordered for companies on petitions under s 994.[211] In the context of 'misconduct' claims, *Anderson v Hogg*[212] and *Bhullar v Bhullar*[213] are illustrations where the petitioners both alleged breaches of duty by the controlling directors of the company. Clearly, derivative claims could be applied in both cases. On the other hand, the case of *Cutland*[214], referred to in Section 4.2, may serve as a typical example of a claim for 'mismanagement'. Where that case triggers the *Chime Corp* criteria, quantifying the damage caused to the company by the wrongful acts of the directors and

202 ibid.
203 *Chime Corp,* (n 198).
204 *Charnley,* (n 144).
205 ibid, at 783.
206 ibid, at 784.
207 S Perera, (n 163) 22.
208 *Chime Corp,* (n 198).
209 ibid.
210 ibid.
211 B Hannigan, 'Drawing boundaries between derivative claims and unfairly prejudicial petitions' (2009) 6 *Journal of Business Law* 606, 624.
212 *Anderson v Hogg* [2002] BCC 923.
213 *Bhullar v Bhullar* [2003] EWCA Civ 424, 2 BCLC 241.
214 *Cutland*, (n 147).

determining the value of liability could be a complex and time-consuming exercise. Nonetheless, Hannigan reasonably countered this point because in *Cutland*, identifying how much money was flowing into the directors' pockets was an essential part of the court's assessment of the affairs of the company at the pleading stage.[215] Following this logic, the step of quantifying directors' liability would not necessarily add to the workload of the courts. Accordingly, the legal framework set out in *Charnley Davies* and *Chime Corp* would provide a potentially desirable approach of drawing the boundary between s 994 and derivative jurisdiction at a relatively small cost.

Besides, the concept of unfair prejudice could be adjusted to take into account the collective interests of all shareholders when the courts assess whether corporate actions in relation to s 994 should continue.[216] In reality, the textual basis for allowing the expansion of the concept is the broad wording of s 994, especially the wording of 'of its members generally'.[217] It is thus feasible to adjust the scope of the interpretation of unfair prejudice to accommodate the broader collective concept.[218] In this regard, Payne persuasively maintains that the ratification and the views of independent bodies within the company are central tools in class proceedings to protect companies from unnecessary litigation.[219] Although these complex collective concepts conflict with the personal nature of traditional unfair prejudice remedies, they are largely relevant if such remedies are to be used to redress corporate wrongs.[220] In this sense, the concept of unfair prejudice with the inclusion of collective interest considerations may be more conducive to maintaining a balance between the interests of shareholders and the company.

(ii) Scenario 2: Using s 994 to Obtain Personal Relief on A Corporate Claim

As previously analysed in Section 4.2.2, the 'no reflective loss' principle may be set aside by the courts when a shareholder seeks personal relief for corporate wrongs based on an unfair prejudice petition. Nevertheless, Section 5.2.2 of this article argues that the basic position of prohibiting unfair prejudice clauses from being a means for shareholders to recover reflective loss should be firmly established to discourage double recovery by defendants, unless two exceptions are involved. As the 'no reflective loss' principle applies strictly to derivative claims, this section will refer to some extent to the case law relating to such claims.

215 B Hannigan, (n 211) 625.
216 J Payne, (n 122) 660.
217 ibid.
218 ibid.
219 ibid.
220 ibid.

Notably, before discussing whether shareholders can use s 994 to seek personal remedies for corporate wrongs, it is necessary to distinguish between those for whom the director is liable.[221] In terms of a s 994 petition, firstly, where the director is liable only to the company, the court should either require the company to bring a corporate claim under *Foss* rule[222] or allow the shareholder to assert the company's rights once the thresholds in Pt 11 of the CA 2006 are met, but limit the shareholder's recovery of reflective loss. Secondly, where the director is liable only to the shareholder, it would be reasonable to grant personal relief. Thirdly, the applicability of the 'no reflection loss' argument should be further analysed in situations where directors may be liable to both the company and the shareholders. Understandably, some complex facts probably blur the boundaries of directors' liability between the company and the shareholders. However, if this distinction is ignored, a shareholder's claim for a breach of his personal rights is likely to be interpreted in general terms as falling within the scope of derivative jurisdiction,[223] which may undermine the availability of s 994.

After clarifying to whom the directors are responsible, the court can scrutinise two situations where the 'no reflection loss' doctrine is breached. The first situation is that the directors are liable for the company's loss but the company lacks a cause of action.[224] In this circumstance, the petitioner's actions are unlikely to reduce the value of the company's assets or to harm the interests of other members.[225] The respondent would also not face a dual claim from both the company and the petitioner. Furthermore, it is important to mention that the mere fact that the company decided not to pursue a claim against the directors is not sufficient grounds for recovery by the shareholders.[226] For instance, in *Giles v Rhind*, the company was unable to afford security for costs due to the serious misconduct of the wrongdoer.[227] That is to say, the severity of the directors' wrongdoing towards the company may be an essential factor in determining whether the company has a cause of action.

The second exception to the application of the 'no reflective loss' principle is that a director is liable to both the company and the shareholder, but the petitioner can prove that his loss is 'separate and distinct' from that of the company.[228] In this regard, the controversial issue is whether the diminution in the value of the petitioner's shareholding resulting from a breach of directors'

221 S Perera, (n 163) 12-13.
222 *Foss*, (n 118).
223 DD Prentice, (n 70) 67.
224 *Johnson v Gore Wood* [2002] 2 AC 1, [2001] 1 All ER 481, at 35.
225 *Lee v Sheard* [1956] 1 QB 195–196.
226 *Gardner v Parker* [2004] EWCA Civ 781, [2005] BCC 46 [49] (Neuberger LJ).
227 *Giles v Rhind* [2002] EWCA Civ 1428.
228 Johnson (n 224) at 35–36.

duties is a personal loss independent of the company's loss.[229] Some past precedents suggest that shareholders have a direct interest in the profits of the company, so shares can be regarded as the personal property of shareholders.[230] In this sense, the reduction in the value of the shares amounts to an impairment of the shareholders' property.[231] Nevertheless, Lord Millett correctly points out that personal losses relating to the value of the shares are usually reflected in the losses of the company, which indicates that it is more appropriate to recover the shareholder's losses through corporate relief.[232] In this way, the shareholder's personal loss is possibly not 'separate and distinct' from the company's loss. In contrast, *Prudential Assurance v Newman Industries (No 2)*[233] is a relevant template. In *Prudential*, the directors had called a meeting by fraudulent circular and the shareholders were allowed to recover any losses suffered as a result.[234] Although the directors' fraudulent conduct also caused a loss to the company,[235] the shareholders' loss could be visibly distinguished from the company's loss and therefore the application of the 'no reflection loss' principle could be excluded. Nonetheless, the shareholders and the company can only recover for the losses they have suffered separately. It should be borne in mind that obtaining two recoveries from the defendant is not tolerated by the principles of equity.

VI. Conclusion

Due to the closely-held and owner-managed nature of private companies, minority shareholders typically expect to enjoy the profits and participate in the management of the company, but they may easily be excluded from management by the majority shareholders using their overwhelming voting power. Therefore, how to enhance the minority shareholder protection in private companies has been a common topic in the study of the unfair prejudice remedy regime. Within the framework of the unfair prejudice remedy, the scope of application under s 994 and the scope of the remedy under s 996 is very broad due to the uncertainty and ambiguity of the wording of the statute itself. Thus, judicial discretion will be a significant means of giving specific meaning to the regime in each case.

Notably, excessive court sympathy for minority shareholders probably results in judicial discretion cutting across the reasonable interests of the majority

[229] S Perera, (n 163) 14–15.
[230] *Bligh v Brent* (1837) 2 Y&C 268, at 408.
[231] JLS Lin, 'Barring recovery for diminution in value of shares on the reflective loss principle' (2007) 66 Cambridge Law Journal 537, 542.
[232] Johnson (n 224) at 66.
[233] Prudential (n 19).
[234] ibid.
[235] ibid.

shareholder and the company, which is likely to adversely affect normal corporate governance. The argument advanced here is that a s 994 petition is not intended to simply protect minority shareholders from any unfair and abusive behaviour by the majority shareholder, but rather to maintain the proper balance of interests of the parties in the company. This requires that the relevant judicial discretion be exercised in a manner consistent with the criteria of 'efficiency' and 'fairness' to maximise the effectiveness of the unfair prejudice remedy system. In general, economically efficient shareholder remedies will minimise the cost to the parties (including the court, the petitioner and the company) of engaging in litigation to achieve redress. In addition, under the guiding standard of fairness, the court will need to be careful to determine whether the protection afforded to minority shareholders by a s 994 petition is likely to swallow up the legitimate interests of majority shareholders. In other words, s 994 cannot be deliberately used as a weapon by devious minority shareholders to break the majority rule.

A review of recent case law in this article reveals that the courts have taken a flexible approach to the interpretation of s 994, which has had a positive impact on protecting minority shareholders' confidence in their investments. However, the pro-minority shareholder judicial approach poses other potential difficulties. In the first place, the overly wide scope of unfair prejudice remedies brings with it extensive fact-finding, which increases the length and unpredictability of such cases. In this way, not only the courts and minority shareholders, but also companies and stakeholders may be drawn into such inefficient proceedings, wasting their time and costs. In the second place, the broad interpretation of s 994 goes beyond traditional personal relief, so shareholders have the opportunity to bring s 994 petitions to get around the substantive ('no reflective loss' principle) and procedural (the leave proceeding and the two-stage proceeding) hurdles in derivative actions. As the s 994 actions lack a reliable threshold for screening frivolous claims and a position prohibiting recovery of reflective loss, they are open to abuse. In this sense, both the effective functioning of the company and the dominance of the majority shareholder may be challenged, which is contrary to the objectives that the unfair prejudice remedy is intended to achieve.

As a consequence, the effectiveness of a s 994 petition and the appropriate judicial discretion are closely linked. While the court's interpretative discretion is a source of vitality for the unfair prejudice remedy regime, there is still a need to clarify its uncertainty to some extent to avoid over-protection of minority shareholders. In order to increase legal certainty and maintain the flexibility of the courts to grant relief to shareholders, there is a need to restructure the judicial discretionary framework concerning s 994 petitions. There are two potential solutions to the above dilemmas. For one thing, the two statutory presumptions proposed by the English Law Commission should be re-adopted to facilitate a more expeditious finding of unfair prejudice and the granting of the

corresponding remedies. This article analyses the critical views of the CLRSG on these reform measures, but finds them unconvincing. For another, when examining a corporate claim on a s 994 petition, the court should divide the petitioner's request into corporate and personal relief. In the case of corporate relief, the court may follow the basic line of interpretation embodied in *Charnley Davies* and *Chime Corp*, and consider collective factors that probably affect the company as a whole prior to a full trial. In terms of personal relief, the court may make appropriate reference to the application of the 'no reflective loss' principle to derivative claims, subject to the categorisation of those for whom the directors are responsible.

Cambridge Law Review (2022) Vol VII, Issue 2, 103–118

A Necessary Shift from Shareholder Primacy toward Stakeholder-Conscious Governance in Light of Corporate Social and Environmental Responsibility

Sophie Treacy*

Abstract

As concerns over climate change continue to loom large in global economic policy, increasing pressure is being mounted on corporate directors to counteract the rapid environmental degradation that is occurring all across the world. The traditional shareholder primacy model of corporate governance, however, fetters the decision-making power of company directors to profit-maximising activities at the expense of other stakeholders, such as customers, employees, and the environment. This inevitably gives rise to a tension between corporate governance norms and sustainable, socially responsible governance. This article argues that, at the level of doctrine, corporate purpose is undergoing a paradigm shift from strictly shareholderist to stakeholder-conscious governance, prompted by a growing number of social and environmental exigencies. The origins and normative legitimacy of shareholder primacy will be explored, along with the extent to which shareholderist governance can be reconciled with activities of corporate social responsibility. It will be submitted that ultimately, shareholder primacy is teetering on the brink of collapse, as the climate crisis demands corporate purpose to evolve toward a much more holistic, stakeholder-conscious model of governance.

Keywords: Corporate Governance; Environmental Social Governance; Corporate Social Responsibility; shareholder primacy; climate change

* BCL Candidate (Oxon), LLB (Trinity College Dublin). I am grateful to the anonymous reviewers for their comments on earlier drafts. Any errors that remain are my own.

I. Introduction

Is the role of the corporation in society to do *well*, or to do *good*? This question neatly encapsulates the ethical quandary that resides at the heart of one of the most enduring and spirited debates in corporate law theory. Despite being the subject of a remarkable body of academic literature produced by generations of leading corporate scholars, the question of what purpose the corporation should serve in society has yet to be met with a definitive answer. In the absence of a sound overarching teleology in respect of corporate purpose, a substantial amount of resulting confusion has permeated corporate law theory. Differing perspectives on whom corporations should fundamentally serve, whether that be its shareholders or wider society, bear a direct impact on how corporations are governed, including the extent to which company directors strive toward more socially responsible governance at the expense of straightforward shareholder profit-maximisation. How the interests of shareholders, stakeholders and wider society are reconciled within company operations never remains static but rather oscillates between shareholderist and stakeholderist orientated paradigms, often spurred on by scandals in corporate governance or times of crisis. Thus, the shareholder primacy norm that is currently said to dominate corporate governance in the United States and United Kingdom is highly susceptible to change. This article argues that the current model of shareholder primacy in corporate governance is on the brink of collapse and is no longer sustainable as the climate emergency, along with many other societal factors, move to the centre of the economic and political agenda. In light of this, it suggests that a shift toward a more "stakeholder-conscious" model of governance could draw corporate law theory into line with reality. Section II profiles the rise of the shareholder primacy norm within corporate governance and draws on empirical studies to demonstrate that whilst it does permit a degree of strategic and profitable endeavours of corporate social responsibility (CSR), it ultimately limits its full implementation. Section III questions the normative power of shareholder primacy in present day corporate governance and examines how greater consideration of stakeholder interests would widen the scope for company directors to engage with CSR. Following from this analysis, it will be submitted that pure shareholder profit maximisation is growing progressively out of touch with corporate governance practice at a time when it is increasingly unacceptable for corporations to simply do well on behalf of their shareholders. They must do good also.

II. The Rise of Shareholder Primacy and its Implications for Corporate Social Responsibility

A. Shareholder primacy: an answer to the agency problem

In their seminal work, *The Modern Corporation and Private Property,* authors Berle and Means identify the emergence of a separation of ownership and control when it comes to how quasi-public corporations are governed.[1] This phenomenon has led to an inevitable problem of agency, as the directors entrusted with the corporation's affairs may have different agendas to that of the shareholders on behalf of whom they act.[2] When ownership and management are "not housed in the same person,"[3] this raises the spectre of a conflict of interest as company directors are endowed with "wide powers"[4] to engage in activities that may ultimately reduce shareholder value.[5] As Bebchuk highlights, without sufficient safeguards, this agency paradigm opens the door to directorial mismanagement of corporate assets, such as self-dealing, excessive pay or the rejection of beneficial acquisitions.[6] The separation of ownership and control places company directors in the driving seat of the corporation, with shareholders sitting passively in the backseat as mere "suppliers of capital,"[7] which gives rise to a significant problem of agency. In response to the centralisation of corporate power around company managers, the concept of corporate governance has presented a solution by fettering the discretion of directors by ascribing to them a plethora of fiduciary obligations in discharging their duties. To this end, corporate governance has been described as "a system to curb the excesses and follies of despotic company bosses,"[8] as it prevents managers from acting in their own interests at the expense of the corporation and its shareholders.[9] Corporate governance rules generally ascribe a number of duties to company directors that work to ringfence their actions around the interests of the company and the shareholders who hold

1 Adolf Berle and Gardiner Means, *The Modern Corporation and Private Property* (Transaction Publishers 1932) 110.

2 Robert Monks and Nell Minow, *Corporate Governance* (4th edn, Wiley India 2010) 94.

3 Harold Demsetz, 'The Structure of Ownership and the Theory of the Firm' (1983) 26 Journal of Law and Economics 375.

4 Blanaid Clarke, 'Corporate Responsibility in Light of the Separation of Ownership and Control' (1997) 19 Dublin University Law Journal 50.

5 Lucian Arye Bebchuk, 'The Case for Increasing Shareholder Power' (2005) 118 Harvard Law Review 833, 843.

6 ibid 843.

7 Hadiye Aslan, 'Shareholders Versus Stakeholders in Investor Activism: Value for Whom?' (2020) Journal of Corporate Finance Volume 60.

8 John Holland, 'The Corporate Governance role of Financial Institutions in their Investee Companies' ACCCA Research Report No 46.

9 Anne Tucker, 'The Citizen Shareholder: Modernizing the Agency Paradigm to Reflect How and Why a Majority of Americans Invest in the Market' (2012) 35 Seattle University Law Review 1299.

"ultimate authority over its business,"[10] due to the otherwise powerless position of shareholders under the agency paradigm.[11]

When it comes to ascertaining whether or not managerial action accords with the best interests of the company and its shareholders, profit maximisation has emerged as a helpful litmus test for corporate directors to employ.[12] Indeed, many commentators have come to treat the maximisation of shareholder profit as synonymous with the exercise of good governance. Robert Clark has conflated the director's fiduciary duty to act in the best interests of the corporation with the maximisation of shareholder wealth,[13] whilst Bainbridge submits that directors should be "obliged to make decisions based solely on the basis of long-term shareholder gain."[14] Such a model of corporate governance that hinges so exclusively on shareholder return can trace its doctrinal foundations back to the decision of the Michigan Supreme Court in *Dodge v. Ford Motor Co.*,[15] where it was noted that corporate activities are conducted "primarily for the profit of the stockholders."[16] As Berger notes, however, the concept of profit maximisation as the cornerstone of the shareholder primacy norm within governance theory was primarily fleshed out in the context of academic discussion.[17] The rise of shareholder profit maximisation as the prevailing norm in company administration can be attributed to the separation of ownership and control within corporations, which necessitates that the wide discretion afforded to company directors be somewhat restrained.

B. Implementing corporate social responsibility under the shareholder primacy paradigm

In accordance with the shareholder primacy norm currently embedded in corporate governance practice in the United States and United Kingdom, it necessarily follows that as a general rule, directors must discharge their duties in a way that maximises shareholder wealth. Whilst activities of corporate social responsibility have historically been portrayed as a natural adversary to the pursuit

[10] Joseph Bower and Lynn Paine, 'The Error at the Heart of Corporate Leadership' (2017) Harvard Business Review 50.

[11] David Yosifon, *Corporate Friction: How Corporate Law Impedes American Progress and What to Do About It* (Cambridge University Press 2018) 10.

[12] Tucker (n 9) 1300.

[13] Robert Clark, *Corporate Law* (Little Brown US 1986) 678.

[14] Stephen Bainbridge, 'Director Primacy: The Means and Ends of Corporate Governance' (2003) 97(2) Northwestern University Law Review 547, 573.

[15] [1919] 170 NW 668, 684.

[16] ibid 499.

[17] David Berger, 'In Search of Lost Time: What If Delaware Had Not Adopted Shareholder Primacy?' in Steven Davidoff Solomon and Randall Stuart Thomas, *The Corporate Contract In Changing Times: Is the Law Keeping Up?* (The University of Chicago Press 2019) 48.

of company profit, a number of empirical studies in recent years have demonstrated that company value maximisation and engagement with CSR are not always mutually exclusive.[18] Indeed, if there is a legitimate "business case"[19] for CSR as a valuable corporate activity in terms of shareholder return, this renders it part and parcel of managerial duties in accordance with the shareholder primacy norm inherent in corporate governance.[20] Such a "business case"[21] for CSR was a prominent feature of the 2005 United Nations Conference, 'Who Cares Wins,' where the concept of Environmental Social Governance (ESG) was first developed, denoting a set of criteria through which socially and environmentally conscious investors can screen the governance standards within corporations.[22] In recent years, the implementation of ESG measures within corporations has been gaining increasing momentum in the minds of institutional investors. This has had an inevitable knock-on effect in terms of how company directors calibrate governance strategies that will result in long-term shareholder value.[23] The causative effect between ESG compliance and corporate revenue has been subject to a number of conflicting hypotheses in terms of the extent to which ESG impacts shareholder return.[24] Nonetheless, there is a strong argument arising from the literature in this area which suggests that the implementation of corporate social responsibility through ESG measures is in fact conducive to long-term wealth maximisation, the ultimate goal of company directors pursuant to shareholder primacy. Before the advent of ESG, Waddock and Graves reported a "significant positive relationship"[25] between the realisation of corporate social responsibility and several key financial performance indicators. More recently, Busch and Bassen have conducted a meta-analysis of over 60 empirical examinations into the nexus between ESG and corporate financial performance, concluding that there is a "clear"[26] positive correlation between the two. A slightly less definitive conclusion was reached in a second-level review produced by Halbritter and Dorfleitner, who

18 Pierre Allegaert 'Codetermination and ESG: Viable Alternatives To Shareholder Primacy?' (2020) 52 International Law and Politics 641.

19 Dorothy Lund and Elizabeth Pollman, 'The Corporate Governance Machine' (2021) 121 Columbia Law Review 2563.

20 ibid 2613.

21 ibid 2613.

22 Gordan Scott 'Environmental, Social, and Governance (ESG) Criteria' <https://www.investopedia.com/terms/e/environmental-social-and-governance-esg-criteria.asp> accessed 1 April 2022.

23 Alexander Kraik, 'Environmental, Social, and Governance Issues: An Altered Shareholder Activist Paradigm' (2020) 44 Vermont Law Review 493.

24 Stuart Gillan, Andrew Koch and Laura Starks, 'Firms and Social Responsibility: A Review of CSR and ESG Research in Corporate Finance' (2021) 66 Journal of Corporate Finance 101889.

25 Sandra Waddock and Samuel Graves, 'The Corporate Social Performance-Financial Performance Link' (1997) 18(4) Strategic Management Journal 303.

26 Gunnar Friede, Timo Busch and Alexander Bassen, 'ESG and Financial Performance: Aggregated Evidence From More Than 2000 Empirical Studies' (2015) 5(4) Journal of Sustainable Finance & Investment 210.

noted a "mostly positive connection"[27] between a company's financial performance and its compliance with environmental and social governance standards. Another meta-study conducted by Deutsche Bank found that corporations with higher ESG ratings outpaced their peers in terms of financial performance.[28] In terms of a single corporation analysis, Ekatah *et al* have investigated the link between corporate social responsibility and company profit based on the annual reports of mega-corporation Royal Dutch Shell Plc, concluding that that CSR is positively related to profitability.[29] Although many commentators have argued that the causative link between corporate social responsibility and the maximisation of shareholder return is "in need of more research"[30] and that the empirical results are often "ambiguous, inconclusive, or contradictory,"[31] the literature in this area certainly debunks any contention that corporate social responsibility and profit maximisation are always at odds with one another.

It is not difficult to identify the reasons why ESG compliance might result in improved shareholder return. For one, several commentators have suggested that directors of companies with good ESG credentials display an overall better quality of management, which is vital to shareholder value.[32] It has also been noted that corporate social responsibility is "highly associated"[33] with good corporate governance and socially responsible firms tend to be the ones with the most efficient management structures, which translates into improved financial output more generally.[34] Moreover, as investor demand for more sustainable business is quickly becoming a "firmly entrenched market reality,"[35] socially and environmentally conscious governance is becoming "critical"[36] if company directors wish to attract new investment. As Williams notes, the investment community is becoming increasingly concerned with whether corporations are implementing environmental stewardship among other socially responsible measures when choosing how to invest.[37] This leaves corporate directors with ample opportunity to attract new socially responsible investments by engaging

27 Gerhard Halbritter and Gregor Dorfleitner, 'The Wages of Social Responsibility - Where Are They? A Critical Review of ESG Investing' (2015) 26 Review of Financial Economics 25, 26.

28 Mark Fulton *et al.,* 'Deutsche Bank, Sustainable Investing: Establishing Long-Term Value and Performance' (2012) SSRN Electronic Journal.

29 Innocent Ekatah *et al* 'The Relationship Between Corporate Social Responsibility and Profitability: The Case of Royal Dutch Shell PLC' (2014) 14(4) Corporate Reputation Review 249.

30 Gillan (n 24) 7.

31 Christophe, Revelli and Jean-Laurent Viviani, 'Financial Performance of Socially Responsible Investing (SRI): What Have We Learned? A Meta-Analysis' (2015) 24(2) Business Ethics: A European Review 158.

32 Oliver Williams, *Corporate Social Responsibility* (Routledge 2014) iv.

33 Yasemin Zengin, 'Corporate social responsibility in times of financial crisis' (2010) 4(4) African Journal of Business Management 382.

34 Theodore Syriopoulos, 'Corporate Social Responsibility and Shareholder Effects: The Greek Paradigm' (2007) 8(1) Journal of International Business and Economy 161.

35 Allegaert (n 18) 672.

36 Lund and Pollman (n 19) 2613.

37 Williams (n 32) iv.

with ESG initiatives, resulting in business growth and ultimate shareholder return. Furthermore, a number of studies have reported the positive effect of CSR in terms of enhancing broader operational drivers of business value, including reputational capital,[38] employee pride,[39] brand differentiation,[40] customer loyalty[41] and improved recruitment[42] to name but a few. One study conducted in collaboration between New York University and the University of Texas investigated the casual effect between corporate philanthropy and company value, reporting a resulting revenue growth in corporations that are sensitive to consumer perception.[43] This reflects the fact that corporations do not operate in isolation from the society around them, but rather in a world of knowledge-based competition, where socially responsible management can yield real effects in terms of the company's bottom line.[44] As Porter and Kramer report, the competitive marketplace within which companies today operate necessitates that they acquire a workforce that is "educated, safe, healthy, decently housed, and motivated by a sense of opportunity."[45] This would suggest that companies have a vested financial interest in improving the communities within which they function, leading to a "convergence of interests"[46] between corporations and wider society. As Galbreath notes, "more than half" of a corporation's assets today are intangible in nature, such as good will, reputation, and human capital.[47] This forces corporate directors to revaluate any position that presents social and economic objectives as distinct and competing, given the wealth of empirical evidence that demonstrates how socially-minded endeavours can also yield economic improvements. Although it is difficult to systematically measure the business benefits of social activity,[48] there is a strong case to be made for an "enlightened self-interest perspective"[49] in terms of how directors, who are fiduciarily mandated to increase shareholder return, implement the requirements of CSR.

38 ibid iv.

39 Arthur Gautier and Anne-Claire Pache, 'Research on Corporate Philanthropy: A Review and Assessment' (2015) 126(3) Journal of Business Ethics 343.

40 Dwane Hal Dean, 'Consumer Perception of Corporate Donations: Effects of Company Reputation for Social Responsibility and Type of Donation' (2004) 32(4) Journal of Advertising 91.

41 ibid 102.

42 Valentinas Navickas and Rima Kontautiene, 'Influence Of Corporate Philanthropy On Economic Performance' (2011) 12(1) Business: Theory and Practice 15.

43 Terence Lim 'Measuring the Value of Corporate Philanthropy: Social Impact, Business Benefits and Investor Returns' (2010) Committee Encouraging Corporate Philanthropy.

44 Michael Porter and Mark Kramer 'The Competitive Advantage of Corporate Philanthropy' (2002) Harvard Business Review <https://hbr.org/2002/12/the-competitive-advantage-of-corporate-philanthropy> accessed 1 April 2022.

45 ibid.

46 ibid.

47 Jeremy Galbreath, 'Twenty-First Century Management Rules: The Management of Relationships as Intangible Assets' (2002) 4(2) Management Decision 116.

48 Lim (n 43) 1.

49 Gautier and Pache (n 39) 7.

It is clear that it is possible for directorial engagement with ESG to result in shareholder profit maximisation, particularly in the context of corporations that rely on reputational capital, institutional investment or strong employee networks. From this perspective, CSR and the fiduciary duties of company administrators to maximise shareholder wealth often appear to go hand in hand, as it appears that "good business is better business."[50] This narrative, however, does not present a fully accurate depiction of how CSR is reconciled with the shareholder primacy norm. As Rampal notes, there is an important distinction to be drawn between ethical and strategic CSR.[51] Whilst the former seeks to implement socially responsible governance policies in the genuine interests of the wider community, the latter only seeks to do so in so far as they advance an overall agenda of capital accumulation. In circumstances where is not possible to frame CSR engagement as a vehicle for ultimate economic gain, the shareholder primacy model of governance does not allow for it. This amounts to a very significant limitation on the implementation of CSR under a regime of shareholder focused governance. Thus, as Post has argued, a shareholderist model of governance can be said to remove ethical reasoning from the picture of company administration because if a decision is legal and profitable, it is ethical so far as shareholder primacy is concerned.[52] The "powerful shareholderist orientation"[53] of corporate governance causes the "marginalisation of corporate social responsibility"[54] in business administration to the same extent that it supports it. Shareholder primacy forces company directors to adopt a blinkered focus on capital accumulation, thereby limiting companies to "profit-seeking units"[55] which only endorse CSR to the extent that it is economically strategic to do so. The idea that corporations should embody a "narrowly self-interested homo economicus"[56] certainly does not sit comfortably with the idea of companies embodying a "corporate conscience"[57] and contributing to the overall quality of life of their workforce, their communities and society at large, even when such a practice that might not necessarily result in improved shareholder return. David Yosifon has commented that this gives rise to a fundamentally uneasy relationship between shareholder focused governance and CSR, as the shareholder primacy paradigm "forces a grinding, rough

50 David Jones, *Who Cares Wins: Why Good Business Is Better Business* (Financial Times Publishing 2011).

51 Anuj Rampal, *Corporate Social Responsibility* (Creative Impact Publishing 2017) 13.

52 Frederick Post, 'The Social Responsibility of Management: A Critique of the Shareholder Paradigm and Defense of Stakeholder Primacy' (2003) 18(2) American Journal of Business 57.

53 Lund and Pollman (n 19) 2562.

54 ibid 2613.

55 Adolph Berle, 'The Impact of the Corporation on Classical Economic Theory' (1965) 79(1) Quarterly Journal of Economics 25.

56 Gary Becker, 'Altruism, Egoism, and Genetic Fitness: Economics and Sociobiology' (1976) 14(3) Journal of Economic Literature 817.

57 Berle (n 55) 36.

relationship between corporations and the society they are meant to serve."[58] The concept of shareholder profit maximisation cuts against any managerial endeavour that uses corporate assets for broader social purposes that are not strategically linked to wealth maximisation,[59] which amounts to a very significant limitation on the implementation of CSR under the paradigm of shareholder primacy.

The "profit-maximizing norm"[60] infused in the duties of corporate directors legitimately arose in response to the great discretion that they are afforded by the separation of ownership and control in company administration. This model of corporate governance has permitted managers to adopt a blinkered focus on capital accumulation and "ignore the interests of the other constituencies."[61] Although a number of empirical studies have advanced compelling evidence in support of a "business case" for CSR, pointing to the many effects it can yield in terms of long-term economic value, genuine and ethical CSR cannot be truly implemented so long as business decisions are conducted in an "amoral vacuum"[62] induced by shareholder primacy. As Sjåfell and Bruner note, the most that can be achieved under shareholder primacy in terms of CSR is a degree of "weak sustainability,"[63] in contrast the kind of "actual sustainability" that is achieved by the genuine ethically-driven socially responsible behaviour of company directors. Corporate administration that is focused solely on profit-maximisation might not completely inhibit CSR, but it does significantly limit how widely it can be implemented by fettering it to endeavours that translate into long-term profit. At a time when society is increasingly demanding that "companies serve a social purpose"[64] beyond mere capital accumulation, it is time to seriously question the dominance of the shareholder primacy norm and the continuing legitimacy of it ringfencing socially responsible governance to that which manifests in shareholder return.

[58] Yosifon (n 11) 4.
[59] Berle (n 55) 25.
[60] Allegaert (n 18) 642.
[61] Post (n 52) 57.
[62] ibid 57.
[63] Beate Sjåfell and Christopher Bruner, *The Cambridge Handbook of Corporate Law, Corporate Governance and Sustainability* (Cambridge University Press 2020).
[64] Larry Fink, 'Larry Fink's 2018 Letter to CEOs: A Sense of Purpose' <https://www.blackrock.com/corporate/investor-relations/2018-larry- fink-ceo-letter> accessed 1 April 2022.

III. The Fall of Shareholder Primacy and its Implications for Corporate Social Responsibility

A. The decline of shareholder primacy in the face of the climate crisis

Due to the pervasiveness of shareholder primacy within corporate governance, it has often been taken for granted in corporate law.[65] It is important to note, however, that such a system is by no means inevitable.[66] In fact, its position as the jewel in the crown of corporate governance theory was hard won over alternative governance paradigms through some of the most influential academic debates in the history of corporate law. In 1932, the infamous Berle-Dodd debate panned out over the pages of the Harvard Law Review, wherein Adolph Berle contended that corporate powers are "at all times exercisable only for the ratable benefit of the shareholders"[67] in contrast with the position of Merrick Dodd, who argued that corporations are "economic institutions which have a social service as well as a profit-making function."[68] This dialogue between Berle and Dodd, which occurred almost a century ago, crystallises the inherent tension between the dominant shareholderist versus stakeholderist governance ideologies that continues to loom large in the present day. Whilst both positions depart from the starting point of restraining managerial power in the face of the agency problem, they quickly reach a fork in the road when it comes to whose interests directorial discretion should be accountable to. The focus on shareholder return as the prevailing norm in corporate governance only secured its definitive status during the 1970s, when it received a series of highly influential endorsements by members of the Chicago School of economists.[69] Milton Freidman's infamous 1970 article, aptly entitled 'The Social Responsibility of Business is to Increase its Profits,'[70] is often credited with carving out the central place for shareholder profit maximisation that currently amounts to the core objective of corporate enterprise.[71] Before the pursuit of shareholder wealth gained such significant traction toward the latter half of the twentieth century,[72] standard corporate governance practice often entailed taking into account a broader pool of stakeholders, including employees,

[65] David Millon, 'Shareholder Primacy in the Classroom after the Financial Crisis' (2013) 8 Journal of Business and Technology Law 191.

[66] Lund and Pollman (n 19) 2628.

[67] Adolph Berle, 'Corporate Powers as Powers in Trust' (1932) 45 Harvard Law Review 1049.

[68] E Merrick Dodd, 'For Whom are Our Corporate Managers Trustees?' (1932) 45 Harvard Law Review 1144, 1148.

[69] Lynn Stout, 'New Thinking on "Shareholder Primacy' (2012) 2(2) Accounting, Economics, and Law, article 4.

[70] Milton Freidman, 'The Social Responsibility of Business is to Increase its Profits' (*New York Times*, 1970).

[71] Michael Jensen and William Meckling, 'Theory of the Firm: Managerial Behavior, Agency Costs, and Ownership Structure' (1976) 3 Journal of Financial Economics 305.

[72] Allegaert (n 18) 642.

customers and the wider community.[73] Thus, as Bower and Paine highlight, the prominent status afforded to shareholders within corporate law theory is a "relatively recent development"[74] and is by no means the only intellectually respectable theory of corporate governance.[75] Indeed, the highly shareholder oriented fiduciary duties of directors in the United States and United Kingdom contrasts with countries in mainland Europe,[76] such as the corporate governance systems of Germany and France, which adopt a more multi-stakeholder approach.[77] This is encapsulated by a recent report of the European Commission, which notes that the managerial role within companies entails balancing the interests of "multiple constituencies."[78] As Yosifon notes, deviations from exclusively shareholder focused governance are "alive and kicking in wealthy, free parts of the world,"[79] demonstrating that purely shareholderist governance only dominates company law theory because we chose to allow it and not because it is the only viable option. Based on the fact that the shareholder primacy norm emerged against the backdrop of much academic debate and is not unanimously applied across the globe, it is clear that it is not the only legitimate governance theory that can be implemented to overcome the agency problem. Rather, the emergence of the shareholder primacy model of corporate governance came about at the expense of an alternative, more stakeholder focused alternative.

Due to the fact that shareholder primacy is not the only feasible solution in terms of transcending the agency problem within corporate governance, its dominance should only persist so long as it can be considered the optimal approach in comparison with other, more stakeholder-centric models of company administration. It is submitted that since the turn of the century, the primacy of straightforward profit-driven governance has begun to erode with increasing vigour. As society is faced with navigating the climate crisis in particular, the tide seems to be turning against the view that shareholder primacy represents the optimal approach to corporate governance. Although in the year 2000, Hansmann and Kraakman famously declared that the rise of shareholder primacy in company administration signalled "the end of history for corporate law,"[80] recent trends in corporate governance indicate that we are on the brink of yet another paradigm shift in corporate law theory when it comes to how shareholder profit

73 Elizabeth Warren, 'Companies Shouldn't Be Accountable Only to Shareholders' (*Wall Street Journal*, 2018).

74 Bower and Paine (n 10) 50.

75 Jensen and Meckling (n 71) 305.

76 John Armour, Simon Deakin and Suzanna Konzelmann, 'Shareholder Primacy and the Trajectory of UK Corporate Governance' (2003) CBR Working Paper Series 1.

77 Martin Gelter, 'Centros, the Freedom of Establishment for Companies and the Court's Accidental Vision for Corporate Law' in Fernanda Nicola and Bill Davies, EU Law Stories (2015) 74.

78 ibid 74.

79 Yosifon (n 11) 171.

80 Henry Hansmann and Reinier Kraakman, 'The End of History for Corporate Law' (2000) Harvard Law School Discussion Paper No 280.

maximisation and CSR are reconciled. As Kaul and Luo observe, public appetite for social responsibility within companies is at an all-time high, as consumers grow increasingly dissatisfied with corporate purpose that is solely devoted to capital accumulation.[81] There is a distinct sentiment emerging from the public sphere that the purpose of business is not only to make a profit, but to foster development and sustainability in wider society,[82] a concept that is gaining increasing amplification across jurisdictions where shareholder primacy has traditionally reigned supreme. Lund and Pollman have recently argued that the "cultural conversation" within society is calling for a "reorientation of corporate purpose away from shareholder primacy"[83] in order to take greater account of broader interests of stakeholders. Stout has detected a "rapid undermining"[84] of the shareholder primacy paradigm in recent years, whilst Hill has characterised shareholder oriented governance as a "one-dimensional model of the past."[85] Such palpable hostility toward traditional shareholderist corporate governance is unsurprising, given that it coincides with a time when the climate crisis represents the defining challenge of our generation.[86] With CSR being lauded as having the potential to drive sustainable development within corporations,[87] the fact that purely profit driven governance partially stifles the implementation of CSR has attracted extensive criticism.

The growing dissatisfaction with traditional shareholderist governance recently came to a head at the American Business Roundtable in 2019, which saw the CEO's of the largest and most influential corporations in the United States renounce blinked shareholderism and commit to "leading their companies for the benefit of all stakeholders."[88] This express endorsement of stakeholder-conscious governance has been heralded as marking a definitive departure from shareholder value maximisation toward more stakeholderist-oriented governance.[89] News reports at the time of the Roundtable describe it as a "turning point"[90] in corporate governance doctrine, which succeeds in overriding "decades of long-held

[81] Aseem Kaul and Jiao Luo, 'An Economic Case for CSR: The Comparative Efficiency of For-Profit Firms in Meeting Consumer Demand for Social Goods' (2018) 39 Strategic Management Journal 1650.

[82] Williams (n 32) 1.

[83] Lund and Pollman (n 19) 2634.

[84] Stout (n 69) 3.

[85] Jennifer Hill, 'Visions and Revisions of the Shareholder' (2000) 48(1) American Journal of Comparative Law 39.

[86] Jacob Poushter and Christine Huang, 'Despite Pandemic, Many Europeans Still See Climate Change as Greatest Threat to Their Countries' (2020) Pew Research Center.

[87] Kristina Herrmann, 'Corporate Social Responsibility and Sustainable Development: The European Union Initiative as a Case Study' (2004) 11 Indiana Journal of Global Legal Studies 205.

[88] Business Roundtable, 'Statement on the Purpose of a Corporation' (2019) <https://opportunity.businessroundtable.org/wp-content/uploads/2019/12/BRT-Statement-on- the-Purpose-of-a-Corporation-with-Signatures.pdf> accessed 1 April 2022.

[89] Lucian Bebchuk and Roberto Tallarita, 'The Illusory Promise of Stakeholder Governance' (2020) 106(1) Cornell Law Review 91.

[90] David Ignatius, 'Corporate Panic About Capitalism Could Be a Turning Point' (*Washington Post*, 2019) <https://www.washingtonpost.com/opinions/even-the-business-moguls-know-its-time-to- reform-capitalism/2019/08/20/95e4de74-c388-11e9-9986-1fb3e4397be4_story.html> accessed 1 April 2022.

corporate orthodoxy."[91] Indeed, soon after the publication of the Business Roundtable statement on the endorsement of stakeholder-conscious governance, the World Economic Forum published a manifesto urging companies to move from "shareholder capitalism to stakeholder capitalism."[92] Larry Fink, the leader of the BlackRock, has recently called all company CEOs to embrace a corporate purpose that serves not only shareholders, but a wide pool of stakeholders, in what he calls "a fundamental reshaping of finance."[93] The resounding message from the Business Roundtable and World Economic Forum is that, as the climate emergency displays no signs of subsiding any time soon, a purely profit driven model of corporate governance teeters on the brink of intellectual collapse.[94] The challenge of climate change forces us to rethink the paradigm that defines the governance norms of the corporations that are causing the greatest amount of environmental harm in our society today, indicating that the time has finally come to dethrone shareholder primacy.[95]

B. IMPLEMENTING CORPORATE SOCIAL RESPONSIBILITY UNDER A "STAKEHOLDER-CONSCIOUS" PARADIGM

It was clear back when the Berle-Dodd debate[96] took place over a century ago that there exists a number of legitimate alternatives to shareholder focused governance that better facilitate the implementation of CSR. Different models of company administration to that of pure profit-maximisation tend to reside on a spectrum depending on much weight they afford to the interests of wider stakeholders including customers, suppliers, local communities and the environment.[97] The greater the significance afforded to the wider pool of stakeholders in directorial decision-making, the greater the scope afforded to company managers in terms of engaging in initiatives that implement CSR requirements. Whilst many commentators have called for reform to the shareholderist model of governance, there has been great deal of discrepancy within the academic literature in terms of the degree to which broader stakeholder interests should encroach upon shareholder primacy. Carl Liao has advocated for

91 David Gelles and David Yaffe-Befany, 'Feeling Heat, C.E.O.s Pledge New Priorities' (*New York Times*, 2019) A1.

92 Jason Karanian, 'And the Winner of the 2020 World Economic Forum is... Stakeholders' (*Quartz*, 2020) <https://qz.com/1791153/winner-of-2020-world-economic-forum- in-davos-stakeholders> accessed 1 April 2022.

93 Larry Fink, 'A Fundamental Reshaping of Finance' (2020) Harvard Law School Forum on Corporate Governance <https://corpgov.law.harvard.edu/2020/01/16/a-fundamental-reshaping-of-finance/> accessed 1 April 2022.

94 Stout (n 69) i.

95 Lund and Pollman (n 19) 2631.

96 See Berle (n 67) and Dodd (n 68).

97 Amanda Wise, 'Corporate Law and the Business Roundtable: Adding to the Debate on Shareholder Primacy vs. Stakeholder Theory' (2021) Corporate Law and the Business Roundtable.

a radical "sweeping overhaul"[98] of the shareholder primacy norm, arguing that it should be replaced with a governance model that takes equal account of all stakeholder interests. Yosifon has also supported an move toward a general "system of multi-stakeholder corporate governance."[99] The majority of commentators, however, have been more conservative in their approaches when suggesting how the shareholder primacy model should evolve in order to pave the way for greater CSR within companies in the age of the climate crisis. David Millon has advanced a theoretical governance framework which appears to be somewhat of a shareholderist-stakeholderist hybrid, which he terms "enlightened shareholder value."[100] This governance model entails corporate directors continuing to pursue shareholder return, but with a more "long-run orientation"[101] that seeks sustainable profits, whilst paying attention to a "full range of relevant stakeholder interests."[102] A similar suggestion has been proposed by Lund and Pollman, who argue for a "reshaping"[103] of shareholder primacy such that it encompasses wider stakeholder interests. They predict that such a paradigm shift is already on the horizon, noting that company administration is increasingly filtered through a more "stakeholder-oriented"[104] lens. Adams and Matheson also seem to have also endorsed a stakeholderist spin on the shareholder primacy norm to allow for greater implementation of CSR, but framed it in the inverse, arguing that company managers should begin with abiding by a stakeholder oriented governance model and then focus on shareholder return later down the road.[105] What certainly emerges from the various suggestions as to how shareholderist governance should evolve to meet CSR-induced demands is that the gap between stakeholder and shareholder governance, which amounted to the great divide borne out in the Berle-Dodd debate, is growing ever smaller.

It is submitted that in the wake of the climate emergency, corporate governance has no choice but to evolve past its current paradigm of straightforward profit maximisation toward a more pluralist management model that allows greater engagement with CSR through taking into account a broader pool of stakeholders.[106] It appears that the route beyond the narrow conception of shareholder primacy, however, does not lie in a sudden radical paradigm shift

[98] Carol Liao, 'Corporate Governance Reform for the 21st Century: A Critical Reassessment of the Shareholder Primacy Model' (2011) 43(2) Ottawa Law Review 187.

[99] Yosifon (n 11) 171.

[100] David Millon, 'Enlightened Shareholder Value, Social Responsibility, and the Redefinition of Corporate Purpose Without Law' (2010) Washington and Lee Public Legal Studies Research Paper Series.

[101] ibid 1.

[102] ibid 1.

[103] Lund and Pollman (n 19) 2615.

[104] ibid 2567.

[105] Edward Adams and John Matheson, 'A Statutory Model for Corporate Constituency Concerns' (2000) 49 Emroy Law Journal 1085, 1086.

[106] Millon (n 100) 3.

toward a multi-stakeholder focused approach. [107] For starters, instituting such a profound and encompassing shift within deeply entrenched corporate governance practice seems highly unrealistic.[108] Rather, it is submitted that company managers should continue operating pursuant to their fiduciary duty to shareholders, but also move incrementally toward governance practices that are "stakeholder-conscious." This would entail directors paying more heed to the interests of stakeholders and the natural environment when discharging their duties in tandem with striving for shareholder wealth accumulation, thereby blurring the line between shareholder and stakeholder oriented governance.[109] Such gradual inclusion of "stakeholder-consciousness" in corporate governance would work to pave the way for the more genuine and holistic implementation of CSR, beyond the "weak sustainability"[110] that is, at best, achieved under the shareholder primacy paradigm. Indeed, this appears to be what is already happening in many global corporations. Writing in the same journal that produced the Berle-Dodd debate, Joly has recently reported that in the present day, "most company leaders believe that their firms' larger purpose is to make a positive difference in the world"[111] beyond simple shareholder wealth maximisation. Indeed, several international mega-corporations have recently released statements that strongly imply the emergence of a more "stakeholder-conscious" and socially aware corporate purpose driving the administration of their businesses. For instance, Google has stated that its purpose is to "organise the world's information"[112] whilst Netflix has claimed that its purpose is to "entertain the world."[113] By framing their corporate around the interests of the "world," it is evident that these leading international companies are moving toward a more outward-looking model of corporate purpose beyond mere profit accumulation. It is clear that the shareholder primacy norm no longer provides a sufficient explanation for why corporations are governed in the manner that they are. The traditional distinction between shareholder and stakeholder focused governance is collapsing at an accelerating rate which can be attributed in particular to the climate emergency that has brought the importance of CSR into sharp focus on the international economic stage. The limitation that shareholder primacy places on the implementation of CSR is no longer acceptable at a time when the climate situation has never been more perilous. It is therefore critical that company law quickly departs from its

107 ibid 21.
108 Lund and Pollman (n 19) 2562.
109 Stout (n 69) 16.
110 Sjåfell and Bruner (n 63) 2.
111 Hubert Joly, 'Creating a Meaningful Corporate Purpose' (2021) Harvard Business Review.
112 Google, 'Our Mission' <https://www.google.com/search/howsearchworks/mission/#:~:text=Maximize%20access%20to%20information,it%20universally%20accessible%20and%20useful> accessed 1 April 2021.
113 Netflix, 'About Netflix' <https://about.netflix.com/en> accessed 1 April 2021.

profit-focused model of corporate governance toward a new, more nuanced and holistic "stakeholder-conscious" paradigm of corporate governance.

IV. Conclusion

As climate change continues to raise international alarm, we must demand more from the corporations that are contributing to environmental degradation. The shareholder primacy paradigm, which once appeared to be the optimum governance solution in response to the agency problem, now seems to be a root cause of environmental destruction. By incentivising a blinkered focus on profit maximisation, the shareholder primacy norm removes other stakeholders, such as the wider community and the natural environment, from the picture of corporate decision-making. To this end, it works to stifle the full implementation of CSR by limiting such endeavours to that which ultimately results in shareholder profit. Although there is an increasing body of empirical evidence that demonstrates how CSR and ESG measures can in many instances coincide with capital accumulation, it is no longer sufficient to limit the implementation of CSR to profit-maximising ventures in the face of the climate crisis. Strict shareholder primacy is on the brink of collapse, as market players from institutional investors to consumers are demanding more socially responsible corporate enterprise. Fortunately, as the Berle-Dodd debate highlighted several decades ago, there are many alternatives to the shareholder focused governance model. It is submitted that in the absence of a major paradigm shift within corporate governance, company directors should increasingly integrate a "stakeholder-conscious" mindset when discharging their fiduciary duties to shareholders. Afterall, for corporate directors, it is no longer a question of simply doing well; the time has come for companies to do good.

Cambridge Law Review (2022) Vol VII, Issue 2, 119–133

The Strange Saga of Compensatory Taxes: Charting a Way Out of India's Maze of Doctrinal Uncertainty

RISHABH JAIN*

ABSTRACT

The 'doctrine of compensatory taxation' ('DCT') – that is, the levy of supposedly non-restrictive taxes by States on account of facilitation of trade, commerce, and intercourse ('TCI') – has been central to India's constitutional jurisprudence on TCI. At the same time, its actual development presents what Chandrachud J rightly calls "a maze of doctrinal uncertainty". This essay closely analyzes the rise and fall of the DCT to reveal underlying conflicts between its two referents which explain the 'maze of doctrinal uncertainty' around it. It then argues that the DCT was rightly rejected, but that the conceptual confusions introduced by it have not been fully extirpated and have even figured in some of the grounds used to reject the DCT. In the process, several key mistakes in Indian constitutional jurisprudence are revealed, including a conflation of two different and contrary doctrines under the common label of DCT, the rejection of the older 'direct and immediate impact' doctrine due to conflation with DCT, and the mislocation and misapplication of the fee-tax distinction. More broadly, the strange saga of the DCT serves as a warning against over-emphasis in TCI jurisprudence on textual factors to the exclusion of conceptual (particularly economic and logistical) ones.

Keywords: Indian Constitution; taxation; trade and commerce; federalism; compensatory taxes

* BA LLB (Hons.) Candidate (National Law University, Delhi). I am grateful to my family and friends for their critical support, to my teachers for their patient instruction, and to the editors and reviewers for their skilful help with this article. Any errors that remain are my own.

I. Introduction

The 'doctrine of compensatory taxation' ('DCT') – that is, the levy of supposedly non-restrictive taxes by States on account of facilitation of trade, commerce, and intercourse ('TCI') – is "a topic that has defined the Indian constitutional experience on inter-State trade, commerce, and intercourse."[1] The purpose of this essay is to assess the nature of the DCT, to assess the implications of these findings for the Indian constitutional jurisprudence on TCI, to evaluate the validity of the DCT in light of such implications, and finally to draw some general analytic and normative conclusions on Indian constitutional jurisprudence on TCI.

Section II of the essay briefly surveys the text and context of the constitutional scheme which gave rise to the controversy. Section III closely analyzes the rise and fall of the DCT to reveal underlying conflicts between its two referents which explain the 'maze of doctrinal uncertainty' around it. Section IV argues that the DCT was rightly rejected, but that the conceptual confusions introduced by it have not been fully extirpated and have even figured in some of the grounds used to reject the DCT. Section V concludes with reflections on the foregoing.

II. The Original Tensions

Part XIII of the Indian Constitution runs from Articles 301 to 307 and regulates 'trade, commerce and intercourse within the territory of India' ('TCI'). Intended to foster the economic unity of India while still carving out space for regional interests in general and the greater needs of underdeveloped regions in particular, it attempts to strike a delicate balance between the powers of the States and the Union.[2] It is thus one of the chief sites for the conflicts that shape federalism in India, and is the origin of the present issue.

Article 301 provides: "Subject to the other provisions of this Part, trade, commerce and intercourse throughout the territory of India shall be free."

Article 304 provides:

> Notwithstanding anything in article 301 or article 303, the Legislature of a State may by law—

1 AP Datar, 'Inter-State Trade, Commerce, and Intercourse' in Sujit Choudhry, Madhav Khosla, and Pratap Bhanu Mehta (eds), *The Oxford Handbook of the Indian Constitution* (Oxford University Press 2016) 489.

2 ibid 488.

(a) impose on goods imported from other States or the Union territories any tax to which similar goods manufactured or produced in that State are subject, so, however, as not to discriminate between goods so imported and goods so manufactured or produced; and

(b) impose such reasonable restrictions on the freedom of trade, commerce or intercourse with or within that State as may be required in the public interest:

Provided that no Bill or amendment for the purposes of clause (b) shall be introduced or moved in the Legislature of a State without the previous sanction of the President.

The competences of the Union and the States with respect to levy of taxes and fees are specified in the Seventh Schedule of the Constitution of India. Significantly for the present purpose, it includes Entry 52, List II ("taxes on the entry of goods into a local area for consumption, use or sale therein"); Entry 57, List II ("Taxes on vehicles … suitable for use on roads […] subject to the provisions of entry 35 of List III ["Mechanically propelled vehicles including the principles on which taxes on such vehicles are to be levied"]"); and Entry 66, List II ("Fees in respect of any of the matters in this List, but not including fees taken in any court").

One more element of the context bears mentioning. S. 92, Commonwealth of Australia Constitution Act ('CACA') provides: "On the imposition of uniform duties of customs, trade, commerce and intercourse among the States, whether by means of internal carriage or ocean navigation, shall be absolutely free." Article 301 was inspired from S. 92, CACA; and, by the time the DCT came to be considered in Indian law, the Australian Courts had settled on two ideas on what would constitute a violation of the freedom of TCI: firstly, the violation would have to be 'direct and immediate'; and, secondly, 'compensatory measures for the purpose of regulating commerce' would not be considered violations of TCI.[3] While Australian TCI jurisprudence has been often referred to by Indian Courts, it has expressly not been dispositive, and has served primarily as a repository of advice rather than as a source of law or precedent.[4] Consequently, it has had a rather unclear and attenuated influence, and shall not be examined here at any length.

With this context, it is now possible to analyze the evolution of the DCT.

[3] ibid 489; *Automobile Transport (Rajasthan) Ltd v State of Rajasthan* MANU/SC/0065/1962 [41] (Das J).

[4] Datar (n 1) 489-492.

III. 'Compensatory Taxation' Before The Courts

A. Baby steps

The Court came to confront this issue first of all in the case of *Atiabari Tea Co Ltd v State of Assam* (1961)[5], which involved a challenge under Article 301 to the Assam Taxation (On Goods Carried by Roads or Inland Waterways) Act, 1954 ('Assam Act'). This law imposed a weight-based tax on tea and jute carried through Assam.[6] All the judges agreed that the construction of the 'freedom' of TCI guaranteed by Article 301 as not including freedom from taxation at all would be unsound on the grounds: that taxes are capable of restricting trade[7]; and that the text of other Articles within the Part, in particular the framing of Articles 304 and the (repealed) Articles 306 as an exception to Article 301, indicates that taxation was within the ambit of Article 301.[8]

Gajendragadkar J (writing for himself, Wanchoo J and Das Gupta J) ('*Atiabari* majority') and Sinha J held that interpreting Article 301 as freedom from all taxation would impermissibly disempower States both against the Union and against the Judiciary.[9] Sinha J also contended that "all taxation is not necessarily an impediment or a restraint" and "may, on the other hand, provide the wherewithals also to improve different kinds of means of transport".[10] On the other hand, Shah J held that Article 301 hits all taxes because "between discriminatory tariffs and trade barriers on the one hand and taxation for raising revenue on commercial intercourse, the difference is one of purpose an[d] not of quality".[11]

The *Atiabari* majority held that it is only "such taxes as directly and immediately restrict trade that would fall within the purview of Art. 301"[12], and Sinha J similarly held that Article 301 protects only "freedom from taxation which has the effect of directly pending the free flow of trade, commerce and intercourse".[13] Finally, the *Atiabari* majority and Shah J struck down the Assam Act as violative of Article 301 and not saved by Article 304(b) due to lack of presidential

5 *Atiabari Tea Co Ltd v. State of Assam* MANU/SC/0030/1960.
6 ibid [22] (Sinha J).
7 ibid [18] (Sinha J), [59] (Gajendragadkar J), [76] (Shah J).
8 ibid [18] (Sinha J), [56]-[58] (Gajendragadkar J), [83]–[85] (Shah J).
9 ibid [60] (Gajendragadkar J), [15] (Sinha J).
10 ibid [16] (Sinha J).
11 ibid [76] (Shah J).
12 ibid [60] (Gajendragadkar J).
13 ibid [19] (Sinha J).

assent.[14] The dissenting opinion by Sinha J upheld the tax as non-violative of Article 301.[15]

Soon though, the Court took a very different position in *Automobile Transport (Rajasthan) Ltd v State of Rajasthan* (1962) ('*Automobile*')[16], which involved a challenge to The Rajasthan Motor Vehicles Taxation Act, 1951 ('Rajasthan Act'), which imposed taxes on vehicles used or owned in the State of Rajasthan, to the extent of denying entry to vehicles not paying such tax.[17]

Das J (writing for himself, Kapur J, and Sarkar J) and Hidayatullah J (writing for himself, Mudholkar J, and Ayyangar J): rejected the view that Article 301 hits all taxes on the ground that the freedom of TCI must be interpreted so as to afford the States reasonable autonomy to carry out their duties and raise finances therefor[18]; and rejected the view that Article 301 hits no taxes on largely the same grounds as the *Atiabari* majority.[19] These two extreme views were also rejected by Subba Rao J, but on the basis of a comprehensive interpretation of Part XIII with a view to harmonious construction.[20]

Thereafter, Das J laid down that "compensatory taxes for the use of trading facilities are not hit by the freedom declared by Article 301" and that

> a working test for deciding whether a tax is compensatory or not is to enquire whether the trades people are having the use of certain facilities for the better conduct of their business and paying not patently much more than what is required for providing the facilities.[21]

On the other hand, Hidayatullah J and Subba Rao J both adverted to the 'direct and immediate impact' test ('DDII') for when a measure, fiscal or otherwise, becomes a restriction.[22] Subba Rao J held that freedom of TCI "is not impeded, but, on the other hand, promoted, by regulations creating conditions for the free movement of trade … with or without compensation."[23] Hidayatullah J held that a levy would have to be more like a fee (that is, *quid pro quo* for some service) than a tax (that is, a levy aimed at revenue generation) to be compensatory.[24]

14 ibid [71] (Gajendragadkar J), [88] (Shah J).
15 ibid [30] (Sinha J).
16 *Automobile* (n 3).
17 ibid [5]-[6] (Das J).
18 ibid [18] (Das J), [186] (Hidayatullah J).
19 ibid [20] (Das J), [190] (Hidayatullah J).
20 ibid [62]-[63] (Subba Rao J).
21 ibid [21] (Das J), [27] (Das J).
22 ibid [54] (Subba Rao J), [186] (Hidayatullah J).
23 ibid [62] (Subba Rao J).
24 ibid [194] (Hidayatullah J).

Ultimately, Das J and Subba Rao J upheld the Rajasthan Act as compensatory or regulatory[25]; whereas Hidayatullah J struck it down as a presidentially unsanctioned restriction because it imposed a 'condition precedent' for the entry of vehicles into the State without any mechanism to ensure that the State obtained a 'fair recompense'.[26]

B. Growing pangs

GK Krishnan v. State of Tamil Nadu (1975) held that "any method of taxation which has a direct bearing upon or connection with the use of the highways is apparently valid", provided that "the very idea of compensatory tax is service more or less commensurate with the tax levied."[27]

However, in *International Tourist Corporation v. State of Haryana* (1981) ('*ITC*'), the Court argued that "If the tax were to be proportionate to the expenditure on regulation and service it would not be a tax but a fee."[28] Thus, given the supposed impracticability of calculating the exact expenditure incurred by the government or the benefits provided to the trader, it held that a levy would be saved by the DCT so long as there existed "a specific, identifiable object behind the levy and a nexus between the subject and the object of the levy".[29] The Court also explicitly allowed for the compensatory tax to be mingled with the other kinds, and to be higher than expenditure or benefits.[30]

In *Maharaja Tourist Service v. State of Gujarat* (1991), the Court relied on *ITC* to hold that "what is necessary [to uphold a tax under the DCT] is existence of a nexus between the subject and the object of the levy and it is not necessary to show that the whole or substantial part of the tax collected is utilised".[31]

In *Bhagatram Rajeevkumar v. Commissioner of Sales Tax, Madhya Pradesh* (1995) ('*Bhagatram*'), the test was elaborated as "substantial or even some link between the tax and the facilities extended to such dealers directly or indirectly", and thus upheld a State tax because "augmentation of their finance would enable them to provide municipal services more efficiently, which would help or ease free flow of trade and commerce".[32] Similarly, in *State Of Bihar v. Bihar Chamber Of Commerce* (1996) ('*BCC*'), the Court held that "some connection", whether direct or indirect,

25 ibid [31] (Das J), [64] (Subba Rao J).

26 ibid [195], [205] (Hidayatullah J).

27 *GK Krishnan v. State of Tamil Nadu* MANU/SC/0315/1974 [22], [29].

28 *International Tourist Corporation v. State of Haryana* MANU/SC/0331/1980 [9].

29 ibid [10].

30 ibid.

31 *Maharaja Tourist Service v. State of Gujarat* MANU/SC/0377/1991.

32 *Bhagatram Rajeevkumar v. Commissioner of Sales Tax, Madhya Pradesh* MANU/SC/0959/1995 [8].

"established between the tax and the trading facilities provided" would be sufficient to characterize the tax as compensatory.[33]

Later, in *Jindal Stainless Ltd v. State of Haryana* (2006) ('*Jindal*-2006'), the Court struck down *BCC* and *Bhagatram* as *per incuriam* the law laid down in *Automobile*[34], although it appears to have missed that the criterion of commensurability was effectively rejected as far back in 1981, in *ITC*.[35]

In *Jindal*-2006, the Court laid out a set of distinctions between taxes and fees proper, with compensatory tax being a hybrid of the two, but closer to the latter. The Court held that: (i) taxes proper are based on "the ability or the capacity of the taxpayer to pay"; whereas fee and compensatory taxes are based on the "principle of equivalence""; (ii) "The main basis of a fee or a compensatory tax is the quantifiable and measurable benefit" whereas taxes proper have no direct, identifiable, measurable benefit; (iii) "A tax can be progressive. However, a fee or a compensatory tax has to be broadly proportional and not progressive."; and (iv) fees are levied on "an individual as such" whereas compensatory taxes are levied on "an individual as a member of a class".[36]

Ultimately, *Jindal*-2006 affirmed the 'working test' for compensatory tax developed in *Automobile* and the doctrine of 'direct and immediate effect' developed in *Atiabari*, and even contended that the former is rooted in the latter.[37] It also placed the burden on the State to show the "quantifiable"/"measurable" benefit provided, either in the legislation or before the Court.[38]

Finally, the question of compensatory taxes, along with several others pertaining to TCI, were referred to a nine-judge bench for authoritative resolution.[39]

C. An inevitable demise

The nine-judge bench in *Jindal Stainless Limited v. State of Haryana* (2016) ('*Jindal*-2016') was convened to deal with four questions, which were framed thus:[40]

1. Can the levy of a non-discriminatory tax per se constitute infraction of Article 301 of the Constitution of India?

[33] *State Of Bihar v. Bihar Chamber Of Commerce* MANU/SC/0959/1995 [12].
[34] *Jindal Stainless Ltd v. State of Haryana* MANU/SC/2085/2006 [45].
[35] ibid [12].
[36] ibid [34]–[38].
[37] ibid [30], [45].
[38] ibid [39].
[39] See *Jaiprakash Associates Ltd v. State Of M.P.* MANU/SC/8437/2008; *Jindal Stainless Ltd v State of Haryana* MANU/SC/0260/2010.
[40] *Jindal Stainless Limited v. State of Haryana* MANU/SC/1475/2016 [9] (Thakur J).

2. If answer to Question No. 1 is in the affirmative, can a tax which is compensatory in nature also fall foul of Article 301 of the Constitution of India?
3. What are the tests for determining whether the tax or levy is compensatory in nature? and
4. Is the Entry Tax levied by the States in the present batch of cases violative of Article 301 of the Constitution and in particular have the impugned State enactments relating to entry tax to be tested with reference to both Articles 304(a) and 304(b) of the Constitution for determining their validity?

However, in any event, even judges who answered the first question in the negative pronounced on the second and third queries (indeed, often before the first query), and thus they can be considered independently.

Chief Justice Thakur speaking for himself, Sikri J, and Khanwilkar J, rejected the theory of compensatory taxes on the grounds that: firstly, all taxes are compensatory in "the broader sense" given that they are "meant to serve larger public good and for running the governmental machinery and providing to the people the facilities essential for civilized living"; second, "the concept of compensatory tax obliterates the distinction between a tax and a fee" given that both integrate the element of *quid pro quo*; and third, the application of compensatory tax theory is infeasible in practice.[41]

Singh J agreed with Thakur J and invalidated the DCT on essentially similar grounds.[42] Ramana J agreed with "the consideration, reasoning and conclusion in the judgment of the learned Chief Justice" with respect to compensatory taxes, and further adds that the Constituent Assembly at no point appears to have contemplated the DCT.[43] Bobde J also concurred with Thakur J on the question of compensatory tax.[44]

Chandrachud J struck down the DCT on the ground that it was "vague and indefinite and has produced a maze of doctrinal uncertainty, if not chaos in constitutional litigation", and further because it could be interpreted to mean that even discriminatory taxes could be upheld so long as apparently 'compensatory'.[45]

Bhushan J invalidated the DCT on the grounds that it aims to be effectively a judicially innovated exception to Article 301, whereas "the exceptions laid down

41 ibid [63]–[65] (Thakur J).
42 ibid [153]–[159] (Singh J).
43 ibid [167.26] (Ramana J).
44 ibid [144]–[152] (Bobde J).
45 ibid [470]–[472] (Chandrachud J).

in the constitutional scheme [from Articles 302 to 306] are self-contained and no new exception can be added by judicial interpretation".[46]

Only Banumathi J upheld the DCT, on the ground that overruling it would prejudice the interests of the States in revenue, while taking issue with the 'nomenclature' of the DCT.[47] She also voted to overrule *Jindal*-2006 and upheld *BCC* and *Bhagatram* as being the correct interpretation of the DCT enunciated in *Automobile*.[48]

By a majority of eight to one, the Court in *Jindal*-2016 rejected the DCT, and this is the position that holds in Indian law today. Yet, as the history told above and analyzed below demonstrates, and as judgments have often noted, it was never a coherent, uncontroversial entity to begin with.

D. *In Memoriam*

In fact, it is submitted that the judiciary had evolved two essentially different doctrines under the common name of 'compensatory taxes'. These may be labelled: D_1, which was conceived in *Automobile* and reached its most thorough formulation in *Jindal*-2006; and D_2, which was conceived in *ITC*, and reached its most thorough formulation in *BCC* and *Bhagatram*.

The function of D_1 was: to enable the State to recover the resources spent by it for the provision of trading facilities; to enable the traders to avoid paying more to the State than necessary for the benefits they collectively obtain; and to enable the judiciary to be able to evaluate challenges to TCI taxes on the basis of an objective quantifiable touchstone. At a practical level: even though it was created in the context of a judgment expanding State powers and upholding a taxation measure, D_1 was also relied upon by the dissenting judges who were striking down that measure; was recovered in *Jindal*-2006 on the basis of the prayer made by the assessee, and against the objection of the States[49]; and was again opposed by the States in *Jindal*-2016.[50]

By contrast, the function of the D_2 was: to enable the State to extract an indefinite amount of taxes from traders for any purpose directly or indirectly "connected" to TCI occurring in its territory; to convince (or coerce) the traders to not challenge taxation measures unless they could somehow show the lack of a link between the tax and facilitation of trade; and to enable the judiciary to give

46 ibid [1008] (Bhushan J).
47 ibid [169(c)] (Banumathi J).
48 ibid.
49 *Jindal*-2006 (n 34) [14].
50 *Jindal*-2016 (n 40) [27].

States vast leeway to levy taxes on TCI without having to formally revisit *Automobile*. At a practical level, it was invoked throughout in judgments that upheld States' levies, and was defended against traders by the States in *Jindal*-2006.[51] Perhaps most revealingly, after the rejection of D_2 in *Jindal*-2006, the States largely did not defend 'the DCT' at all in *Jindal*-2016.[52]

At both theoretical and practical levels, then, D_1 and D_2 are essentially different, contrary doctrines; and the "maze of doctrinal uncertainty" that Chandrachud J alluded to is explained in major part as arising out of judicial treatment as interpretations of the same entity. This manifests at the most direct level in the Court evaluating D_1 and D_2 together; evaluating D_1 primarily on its merits but D_2 on its fit with D_1; and, in *Jindal*-2016, striking down 'the DCT' partly because of the confusion or perceived unworkability caused due to a conflation of the two. However, as shown below, there are also other important (but as yet unnoticed) ways that the contest between D_1 and D_2 has shaped -- and confused -- Indian TCI jurisprudence.

At the same time, the *Jindal*-2016 rejection of the DCT requires critical scrutiny *per se*, especially given that it overrules decades of jurisprudence. Further, if the DCT is indeed indefensible, the constitutional distortions introduced to sustain it in Indian law need also to be detected and removed, including where they have been deployed in critiques of the DCT. Finally, it is fair to question whether either or both of D_1 or D_2 could stand or fall on its merits in an unclouded assessment.

IV. Idea(s) Rightly Rejected

A. To be fair

It is now possible to see that some of the criticisms adduced against D_1 and D_2 are incorrect or insufficient to reject them *in toto*. For instance, although *Bhagatram* and *BCC* are indeed *per incuriam*, D_2 could have been considered on its own merits, especially in *Jindal*-2016. Just so, D_1 cannot be held liable for 'doctrinal uncertainty' as such, since it was due to later judicial activities that such confusion arose.

The argument against D_1 that it is really a 'fee' rather than a 'tax' is also irrelevant.[53] As seen above, this argument emerged in the context of the *ITC* Court attempting to argue for D_2 against the obstacle posed by the *Automobile* 'working

51 *Jindal*-2006 (n 34) [20]–[21].
52 *Jindal*-2016 (n 40) [27].
53 For a broader overview of the interplay between the three, see Neha Pathakji, 'Slippery Slopes of Compensatory Tax and Fee' (2014) 56(1) Journal of the Indian Law Institute 78–94.

test'. It perhaps appeared to make sense due to the States' contentions in many cases that they were enacting 'taxes' under List II. However, the two invocations of 'tax' are basically unrelated – the so-called 'doctrine of compensatory tax' is purely about the scope of Article 301 of the constitution, whereas 'taxes' for the purpose of List II are about States' competences in respect of actions that are already constitutional under Part XIII. A fee could be tested on the touchstone of being 'compensatory' just as a tax could be.[54] If the DCT were differently termed (for instance, 'compensatory levy'), the 'fee, not tax' charge would be defused, suggesting that it is devoid of substance.

The originalist arguments adduced by Bhushan J and Ramana J in *Jindal*-2016 against D_1 and D_2 can also not be assigned too much weight. Originalism, in whichever form, has never been dispositive in Indian constitutional jurisprudence, and in fact has been critiqued often by the Court.[55] Moreover, given that the theory of compensatory taxation was in vogue in Australian law when much of the language of Article 301 of the Indian constitution was borrowed therefrom, a reasonable reading of the 'original meaning' of Article 301 could include the DCT.[56]

Then there is the conflation in *Jindal*-2016 of the DDII with the DCT by a majority of the Court.[57] It is true that a majority of judges in the *Automobile* case seemed to subscribe to both the DCT and the DDII, and that *Jindal*-2006 thought that the former was a consequence of the latter. Actually, there is no reason to believe so, given that: compensatory taxes, even in the narrow D_1 sense, may well have a direct and immediate adverse impact on trade; and, just so, many taxes which have only a remote impact may not be compensatory at all. It has even been argued that *Automobile* "practically overruled" *Atiabari*, particularly in respect of the DDII.[58] In fact, it is hard to see how the *Jindal*-2016 view could survive without incorporating a version of the DDII, given that some non-discriminatory taxes surely would be so prohibitive that they would constitute restrictions on trade. The DDII need not necessarily be equated with the stringent approach taken by the *Atiabari* majority either, since it was also relied upon by Sinha J, who endorsed a more expansive view of States' power.

Chandrachud J's argument in *Jindal*-2016 that the DCT legitimates certain types of discriminatory taxation is correct only to a very limited extent, since it is possible to read a non-discrimination criterion into both D_1 and D_2 without changing the essence of either. Moreover, it is hypothetically not altogether

[54] See, for example, *Automobile* (n 3) [47] (Subba Rao J).

[55] See, for example, *Automobile* (n 3) [16] (Das J).

[56] See *Automobile* (n 3) [87] (Hidayatullah J).

[57] See *Jindal*-2016 (n 40), Order [5]–[6].

[58] *Krishnan* (n 27) [12]–[13].

unreasonable to speculate that some variant of differential DCT could actually shed light on the proper construction of 'discrimination', by providing that outstate traders may be made to pay a little more for services availed given that they do not contribute in the capacity of residents.

Thus, a sharper scrutiny not only considerably problematizes the judgment in *Jindal*-2016, but also invalidates many of the arguments adduced against the DCT. However, as shown below, it also leaves many others standing and reveals new ones.

B. Shared flaws

As seen above, D_1 and D_2 both have been justified on the bases: that they do not restrict TCI but merely facilitate it; that, in their absence, States would be effectively disabled from exercising their sovereign/constitutional powers to impose taxes upon TCI; and that Australian constitutional law, from which Article 301 of the Indian Constitution is inspired, also has the doctrine of compensatory taxes.

The third of these arguments is too weak to adduce any real weight. Australian constitutional law in respect of TCI is significantly removed from its Indian counterpart in both text and context. S. 92, CACA refers to TCI being "absolutely free" and provides for no list of restrictions akin to Articles 302-305.[59] CACA was adopted in the context of provinces uniting into a federation whereas the Indian Constitution was adopted in the context of a strong Central government delegating more power to weak units.[60] Finally, in any event, the doctrine of compensatory tax has now been rejected in Australian law.[61]

The second argument is not correct. The *Jindal*-2016 majority view, the DDII, or some permissive combination of the two could easily lead to an equally or more permissive view of States' powers.

The first argument is also incorrect. The invocation of 'trade' as a homogenous activity and 'traders' as a homogenous beneficiary class masks actual disparity in the impact of tax on traders and trade. The facilities provided to all traders are the same, but traders with lower absolute profit margins to begin with are bound to be disincentivized due to higher opportunity costs, whereas those with higher absolute profit margins would actually be partly incentivized due to lowered competition. Further, unless there is rational differentiation between

59 Datar (n 1).
60 ibid.
61 *Jindal*-2016 (n 40) [97]–[105] (Thakur J).

classes of goods, some types of trade are disincentivized more than others, and those with very low profit margins may be discouraged altogether. The point here is not that taxes which restrict trade are problematic *per se*. It is that there is no contradiction between a tax being compensatory and restrictive at the same time – that is, the DCT does not perform its central goal of determining when the freedom of TCI has not been breached by taxation.

C. DISTINCT ERRORS

Then, there are certain conceptually fatal issues which afflict D_1 or D_2 specifically.

Perhaps the central argument for D_1 is that, even if it may have an uneven or 'regressive' impact on traders, it still allows the State to collect its rightful expenditure on "certain facilities for the better conduct of their business". This is submitted to be incorrect. Typically, the concept of facilities in question has been used to evoke the construction and maintenance of highways, waterways, and so forth. *Jindal*-2006 even required governments to show that their provision of trading facilities is quantifiably equivalent to tax levied. In fact, even a first-order assessment, counting only the essentials required to make any TCI through the State possible, would have to include law-and-order, medical facilities for drivers, repair shops for vehicles, and so forth. A more comprehensive assessment would likely reveal that practically any non-wasteful spending, even if purely 'domestic', increases the stability and prosperity of the State, both of which reliably stimulate TCI. Thus D_1 is unfair not only to some traders but also even to the States.

On the other hand, D_2 is essentially impracticable. D_2 may be interpreted in at least two ways, which may be called the weighted and unweighted versions respectively. The latter, which appears to be the one that the Courts actually used, would legitimate tax extraction on the ground that it is of use that such tax is contributing to the State's funds, some of which are going towards the facilitation of TCI. In practice, it is hard to see how any tax could be held invalid on such touchstone. On the other hand, a potential more sophisticated version of D_2 would legitimate extraction of an amount of tax roughly equivalent to the sum of State expenditures each weighted by its respective contribution to the facilitation of TCI. This would be theoretically sound, but likely impossible to put into practice on account of the sheer amount of required foreknowledge.

D. FINAL THOUGHTS

At this juncture, it is pertinent to explore some of the important aspects of the upshot of this inquiry with respect to the change in the Indian structure of taxation powers wrought by the GST Amendment[62], which, among other things, deletes Entry 52, List II (entry taxes).[63]

The GST Amendment is prospective, and thus the (critical, partial) support here to the *Jindal*-2016 judgment consequently supports the State collection of arrears amounting to INR 300 billion in entry taxes.[64] It also rules out the 'compensatory' framing of other frequently levied taxes on TCI, such as, for instance, the vehicles tax under Entry 57, List II.

There are also some lessons which can be inferred here as regards the operation of the GST paradigm. Firstly, the conceptual unworkability of compensatory levy must give pause to arguments grounded in the DCT for 'compensatory cess' in GST.[65] Secondly, the disentanglement of the DCT and the DDII opens the possibility of validly subjecting GST to the latter. Finally, the original causes of the judicial innovation of the DCT—the possibility of tax-caused trade barriers and unclarity on the extent of States' powers of taxation—may still persist in the GST regime[66], so that avoiding the problematic lines of theorization identified in this study remains potentially as important as ever.

V. CONCLUSION

The development of the DCT was a profound judicial misadventure, and attempting to sustain it muddled much of the Indian constitutional jurisprudence on TCI. The attempt here has been to not only argue for the invalidity of the DCT, but also to do so from a perspective extricated from the labyrinthine conceptual confusions introduced to keep the DCT afloat. These include, notably: the conflation of the DCT and the DDII; the conflation of D_1 and D_2; and the mislocation of the 'fee'-'tax' controversy in Article 301. What remains is a fatal critique of the DCT based on its conceptual unworkability, which includes not only legal but also necessarily economic and logistic aspects. In general, as evidenced

62 The Constitution (One Hundred and First Amendment) Act 2016.

63 ibid s 17(b).

64 Ashpreet Sethi, 'The Tax Case That May Cost India Inc Rs 30,000 Crore' (*Bloomberg Quint*, 26 July 2016) <https://www.bloombergquint.com/business/the-tax-case-that-may-cost-india-inc-rs-30000-crore> accessed 24 April 2022.

65 See, for example, Yash Sinha, 'GST Compensation to States: An Ineluctable Obligation on the Union' (2021) 14 NUJS Law Review 17–19.

66 Sethi (n 64).

by this study, the neglect of such aspects is especially hazardous to the interpretation of TCI-related provisions of the constitution, where law in its inderdeterminacy may legitimate any answer, but material reality will brook far fewer.

Cambridge Law Review (2022) Vol VII, Issue 2, 134–146

Justice Shortchanged? Redrawing the Ethical Boundaries of Lifted Judgments Following *Crinion v IG Markets Ltd* [2013] EWCA Civ 587

SHEN-WAY CHONG*

ABSTRACT

In the course of their work judges quite often expose themselves to criticisms. Implicit in these critiques is the expectation that responsible judicial writing should not only be encouraged but required as an ethical obligation. In composing their own legal analysis, judges are expected to have critically deliberated upon the issues in dispute from a neutral perspective. Where a judgment significantly replicates the prose of one litigant, regardless of whether the source is acknowledged, the other litigant would be justifiably indignant with the failure of the judge to devote sufficient thought to his position, notwithstanding the merits of said judgment. This paper will attempt to make the case for the imposition of a duty on the bench not to plagiarise the language of the victorious litigant and to advocate for a "functional approach" to deal with instances of unbridled judicial copying.

Keywords: judicial plagiarism; judgment writing; duty to provide reasons; judicial ethics; functional approach

* Barrister-at-Law (Gray's Inn), LLB(University of Liverpool). I am grateful for the helpful comments from Lady Justice Geraldine Andrews DBE and an anonymous referee on a draft of this article. All omissions and shortcomings remain min.

I. Introduction

We skim off the cream of other men's wits,
pick the choice flowers of their tilted gardens
to set out our own sterile pots.
Robert Burton[1]

From the annals of the seventeenth century, adjudication by impartial and independent judges has been recognized as one of the cornerstones of civil society.[2] Unlike the medieval epoch where the monarch could summon a judge to compel him to account for his actions,[3] judges of the present demonstrate their independent thought process through their judgments. A judgment goes beyond the performance of a bureaucratic function – the supply of judicial reasons is essential to "the establishment of fixed intelligible rules and for the development of law as a science".[4]

When a huge portion of partisan submission is "lifted" to form the crux of a judge's reasons, allegations of bias and partiality will inexorably surface. A lifted judgment is broadly understood as the incorporation of the submissions of one party in whole or in part as the judge's own reasoning without addressing the central arguments raised by the other party or explaining why those arguments were rejected.[5] The audience that a judgment serves needs to understand how the judge analysed the factual circumstances of the case and applied the law accordingly. A lifted judgment either defeats or diminishes these expectations.

Part II of this Article discusses the purpose of judgment writing and the judicial duty to provide reasons. This segment goes further to explore how the courts have dealt with judgments that divulge little to no judicial reasoning, otherwise known as non-speaking judgments.[6] Part III of this Article attempts to illustrate the paradox between lifted judgments and the ethics espoused in the Guide to Judicial Conduct[7] with extensive reference to the leading Court of Appeal decision in *Crinion v IG Markets*.[8] This segment points out the flaws in

1 Robert Burton, *The Anatomy of Melancholy* (Floyd Dell and Paul Jordan-Smith eds, Tudor 1948) 18.

2 John Locke, *Two Treatises of Government*, Book II (Black Swan, 1690) at para 4. On the discussion of Locke's theory, see Peter H. Russell, *The Third Branch of Government* (McGraw-Hill Ryerson, 1987) 20–21.

3 *O'Reilly v Mackman* [1983] 2 AC 237, 252.

4 Herbert Broom, *Constitutional Law Viewed in Relation to Common Law, and Exemplified by Cases* (2nd edn, Maxwell 1885) 147–148.

5 See generally, *Williams v Solicitors Regulation Authority* [2017] EWHC 2005; *Crinion v IG Markets Ltd* [2013] EWCA Civ 587; [2013] C. P. Rep. 41; *English v Royal Mail Group Ltd* (2008), UKEAT/0027/08.

6 *Joinery Plus Ltd (in administration) v Laing Ltd* (2003) 87 Con LR 87; *Soleimany v Soleimany* [1999] QB 785.

7 Courts and Tribunals Judiciary, 'Guide to Judicial Conduct' (March 2020) <https://www.judiciary.uk/wp-content/uploads/2020/03/Guide-to-Judicial-Conduct-Guide-Fourth-Amendment-2020-v3-1.pdf> accessed 20 June 2021.

8 [2013] EWCA Civ 587; [2013] CP Rep 41.

Crinion and argues that it is irreconcilable with the key principles enumerated in the Guide. It concludes with the assertion that a lifted judgment is unethical and poses serious detriment to public confidence in the judiciary. Part IV of this Article calls for a "functional approach" to be preferred over the minimalist prose endorsed in *Crinion* when dealing with instances of extensive judicial copying and proposes an addition to the Guide to clarify the permissible boundaries of judicial copying.

II. Why Write Judgments

To a great extent, the common law has evolved out of swashbuckling advocacy and at the expense of litigants, rankling courtroom dramas.[9] As the stage is set for a contest of averments, the advocates representing their respective clients inject every available strand of learning, suave, persuasion and sometimes emotion in their painstaking attempts to weave an ironclad case. All of these unfold before the bench – justices who are bestowed with the mandate to resolve the dispute by seeking for answers necessary to justify and achieve a fair outcome. Upon hearing arguments from both sides and admitting all relevant evidence, a decision is eventually rendered in the form of a written judgment that encapsulates much of the legal discourse and signifies the culmination of judicial deliberation.[10]

An authoritative judgment can only be rendered if its author is mindful of the purposes of writing.[11] The judgment that entails after a usually protracted dispute is etched into the institutional memory of the court and becomes the law. At the heart of the English legal system is the adversarial nature of proceedings that has existed since time immemorial, and judgments are written to apprise the litigants of "who has won and why".[12] In addition to being a healthy discipline for those who exercise powers that are capable of causing vast harm to others,[13] a written judgment is a vehicle by which the judiciary elucidates, expounds upon and creates rights for the citizenry.[14]

Many arguments can be advanced in support of judgment writing, the most obvious reason being the practice of law is traditionally grounded in literary works. "Most law professors, judges and practicing lawyers devote considerable effort to researching the law and composing a variety of legal writings, including law

9 John D. Heydon, 'Threats to Judicial Independence: The Enemy Within' (2013) 129 LQR 205, 213.

10 *Re B (A Minor)* [1990] 1 FLR 344, 347.

11 James Wilson and Alexander Horne, 'Judgment Matters' (2010) 160 (7446) New Law Journal <https://www.newlawjournal.co.uk/content/judgment-matters> accessed 11 June 2021.

12 ibid. Also see *Meek v City of Birmingham District Council* [1987] IRLR 250; *R v Knightsbridge Crown Court Ex p International Sporting Club (London) Ltd* [1982] QB 304; *R v Harrow Crown Court Ex p. Dave* [1994] 1 WLR 98.

13 ibid 9, at 211.

14 Gerald Lebovits, Alifya V Curtin and Lisa Solomon, 'Ethical Judicial Opinion Writing' (2008) 21 Georgetown Journal of Legal Ethics 237, 244.

journal articles, client memoranda, appellate briefs and legal opinions".[15] The legal significance of judgment writing lies in the coalescence of a variety of sources ranging from knowledge of various factual and legal issues to the proper application of primary and secondary legislation, precedents, evidential matters and other pertinent sources of authority.[16] One only has to take a cursory glance at the contents of a judgment to realise that these are indispensable literatures in the academy of legal writing.

Written judgments have been described as necessary to "constrain judges and promote accountability in the resolution of real world disputes".[17] Writing a judgment impels the judge to exert intellectual discipline on themselves, which in turn reinforces judicial deliberation to the effect that the ultimate decision is derived from reasoned judgment and thoughtful analysis over an arbitrary exercise of judicial authority.[18] For those aggrieved by the trial court's decision and seek to contest those findings before an appellate court, a meaningful appellate review can only be established if the first instance judgment discloses salient grounds of appeal to enable the higher courts to clarify certain issues of law which may yet remain unclear.[19] Although no one savours the prospect of being shown to have erred,[20] the "disinterested application of known law"[21] necessitates the removal of injustice alleged by the contender.[22]

Interwoven with the crafting of judgments is the judicial duty to give reasons for such duty forms the "building blocks of the reasoned judicial process".[23] As the apothegm goes, a judge should not speak but his judgment should.[24] While a judgment is primarily written for the litigants, it does not follow

15 Carol M. Bast and Linda B. Samuels, 'Plagiarism and Legal scholarship in the Age of Information Sharing: The Need for Intellectual Honesty' (2008) 57 Cath. U. L. Rev. 777, 793.

16 Naida Haxton, 'Editing Judgments: Lessons Learned in Law Reporting' (2007) 57 Clarity <https://minio2.123dok.com/dt02original/123dok_es/original/2019/01_24/vrqr1m1579237240.pdf> accessed 11 June 2021.

17 Jeffrey L. Dunoff and Mark A. Pollack, 'Experimenting with International Law' (2017) 28(4) EJIL 1317, 1338. Similarly, see Roman N Komar, *Reasons for Judgment* (Butterworths 1980) 9. Cf Alfred Denning, *Freedom Under the Law* (Stevens & Sons 1949) 91–92.

18 See generally Karl N. Llewellyn, *The Common Law Tradition: Deciding Appeals* (Little, Brown & Co 1960) 13, 26, 56.

19 Andrew Bainham, 'Judgment: Whose Responsibility is it?' (2019) Fam Law 849, 851 (note).

20 ibid. Also see *Re L-B (Reversal of Judgment)* [2013] UKSC 8, [2013] 2 FLR 859 at [46] where Baroness Hale commented on the situation where a judge recognises that an error has been made: "it takes courage and intellectual honesty to admit one's mistakes".

21 A phrase coined by Louis L. Jaffe, *English and American Judges as Lawmakers* (Oxford University Press 1969) 13 when describing the function of a judge.

22 Lord Devlin, 'Judges and Lawmakers' (1976) 39 MLR 1, 3.

23 *Glicksman v Redbridge Healthcare NHS Trust* [2001] EWCA Civ 1097, (2001) 63 BMLR 109 at [11]. It is only a general rule that reasons need to be given. For exceptions to the general rule, see, for example, *Capital and Suburban Properties Ltd v Swycher* (1976) Ch. 319; *Antaios Compania Naviera SA v Salen Rederierna AB* [1985] 1 AC 191 and *R v Harrow Crown Court Ex p. Dave* [1994] 1 WLR 98.

24 Judges Matter, 'Judges Speak Through Their Judgments' (23 September 2019) <https://www.judgesmatter.co.za/opinions/judges-speak-through-their-judgments/> accessed 13 August 2021. In a 2012 lecture, Lord Neuberger stated that: "Without judgement there would be no justice." See Lord Neuberger, No Judgment – No Justice (First annual BAILII Lecture, 20 November 2012) <https://www.bailii.org/bailii/lecture/01.pdf> accessed 13 August 2021.

that the litigants are the sole consumers of the judgment. Judges also write for the public and for professionals including other judges, lawyers, academic scholars and law students.[25] At the broadest level of public accountability, a requirement that judges give reasons for their decisions – grounds that can be debated, attacked and defended – is fundamental to the legitimacy of the judicial institution in the eyes of the public.[26] As Lord Denning explained:

> [I]n order that a trial should be fair, it is necessary, not only that a correct decision should be reached, but also that it should be seen to be based on reason; and that can only be seen, if the judge himself states his reasons.[27]

Recent authorities from the Court of Appeal have equated a non-speaking judgment as an error of law that warrants appellate intervention.[28] Illustrating the sanctity of the judicial duty to provide reasons, Neill LJ in *Re L (Minors)* went as far as holding that a non-speaking judgment was "defective" and ordered a retrial of the matter before another judge,[29] as did Males LJ in *Simetra Global Assets v Ikon Finance*.[30] The exigency that the modern courts have placed on the expression of intellectual substrate and the disdain for reticence mark a tectonic shift from the stance of their predecessors especially if one recalls that Lord Mansfield, at one time was audacious enough to advise: "[N]ever give your reasons; – for your judgment will probably be right, but your reasons will certainly be wrong."[31]

III. Lifted Judgments—A Stain on Judicial Ethics

Most judges strive to be fair and appear to be fair so that the litigants walk away satisfied that they were fully heard, their positions were fully considered, and the pertinent rules were applied properly throughout the proceedings.[32] These considerations run like a golden thread throughout the legal system and is

25 Mary Kate Kearney, 'The Proprietary of Poetry in Judicial Opinions' (2003) 12 Widener Journal of Public Law 597, 601.

26 David L. Shapiro, 'In Defense of Judicial Candor' (1987) 100(4) Harvard Law Review 731, 737.

27 Alfred Denning, *The Road to Justice* (Stevens & Sons 1955) 29.

28 *Simetra Global Assets Ltd v Ikon Finance Ltd* [2019] 4 WLR 112 at [39]; *English v Emery Reimbold & Strick Ltd* [2002] 1 WLR 2409 at [19]; *Flannery v Halifax Estate Agencies Ltd* [2000] 1 WLR 377, 381–382.

29 (CA, 30 January 1991).

30 ibid 28 at [8]. It must however be stated that a retrial is an expensive step in the judicial process and is rarely granted, see *Lai Wee Lian v Singapore Bus Ltd* [1984] 1 AC 729 at 741: "Thus, if the only conclusion open on the evidence trial was the conclusion reached by the trial judge, then, notwithstanding an inadequate statement of reasons, the matter need not go to a new trial."

31 John Campbell, *The Lives of the Chief Justices*, vol 3 (James Cockcroft & Co 1873) 481.

32 This point was more eloquently stated by Sir Robert Megarry: "One of the important duties of the courts is to send away defeated litigants who feel no justifiable sense of injustice in the judicial process. See Robert Megarry, 'Temptations of the Bench' (1998) 16 Alta. L. Rev. 406, 410.

especially discernible in the first instance courts where interaction between judges and litigants occur more readily. [33] It is easy to overlook them when a judge becomes unduly immersed in the proceedings before them or have made up their mind prior to full argument from counsel,[34] for why else an author would deem judicial neutrality as a myth?[35] The scale of such anxiety, though unlikely to be widespread, wields considerable persuasion especially when a judgment is 'lifted' verbatim from a partisan submission while neglecting the core arguments of the other. Understandably, such instance would engender apprehensions that the judge was an unreliable agent of justice who could not be trusted to carry out their constitutional obligation *erga omnes*.

A lifted judgment can be aptly described as judicial plagiarism.[36] Bereft of the judge's independent analysis and contribution, a judgment may appear to be skewed in favour of the party whose submissions were adopted or may at the very least suggest that the judge's mind was shut to the arguments of the losing party.[37] Even if the lifted judgment represents the judge's true thinking, it reflects poorly on the administration of justice.[38] The Guide to Judicial Conduct underscores three distilled principles from the Bangalore Principles of Judicial Conduct[39] that form the essence of judicial ethics – judicial independence, impartiality and integrity.[40] In particular, a judge is expected to display "intellectual honesty"[41] and to "avoid situations which might reasonably reduce respect for judicial office or might cast doubt upon their judicial impartiality".[42] It is difficult to see how a judge who plagiarises the winning submission without any independent thought can be said to have upheld their "Hippocratic" oath of ethics. To argue otherwise would be tantamount to blowing hot and cold.

However, the Court of Appeal in *Crinion v IG Markets Ltd* was reluctant to accept that plagiarism and judicial ethics are mutually exclusive.[43] The gist of the dispute is one of enforceability of debt, yet it is unlikely to be overly significant from a commercial vantage. In a nutshell, *Crinion* is one of the few ironies in

33 Simon Stern, 'Copyright Originality and Judicial Originality' (2013) 63 UTLJ 385, 386.

34 Lord Diplock, for example, was characterised as someone who prepared for oral hearings very thoroughly to the extent that it was not unusual for him to have made up his mind before a hearing. See Alan Paterson 'Does Advocacy Matter in the Lords?' in James Lee (ed), *From House of Lords to Supreme Court Judges, Jurists and the Process of Judging* (Hart Publishing 2011) 257. At p.258, Lord Hope, recalling his days as a barrister appearing before Lord Diplock, said: "He didn't allow arguments to develop that he thought had nothing in them … and really cut you short."

35 Kathleen E. Mahoney, 'The Myth of Judicial Neutrality' (1996) 32 Willamette L. Rev. 785, 788.

36 *Williams* (n 5); *Re S* [2015] EWCA Civ 1015; [2016] 2 FLR 965; *Crinion* (n 5); *English* (n 5).

37 ibid 33, at 393.

38 Per Sir Stephen Sedley in *Crinion* (n 5) at [39].

39 United Nations, 'The Bangalore Principles of Judicial Conduct' (2002) <https://www.unodc.org/pdf/crime/corruption/judicial_group/Bangalore_principles.pdf> accessed 20 June 2021.

40 ibid 7.

41 ibid.

42 ibid.

43 *Crinion* (n 5).

English common law that starts off being about one thing only to end up being about quite another – the fashion in which the trial judge opted to draft his reasons. While much displeasure was expressed against the etiquette of "cut-and-paste",[44] the court ultimately affirmed that a minimalist judgment was neither defective nor capable of giving rise to injustice that justifies appellate intervention.[45] The test formulated by Underhill LJ was "whether the judge properly addressed" the contentions of the losing side,[46] without further explanation or illustration on judicial propriety. The judgment rendered by His Lordship, read in entirety, suggests that so long as the judge provides a brief analysis to show that the conclusions derived were not the product of a purely mechanical act, the judge would have discharged his or her judicial duty to provide reasons, notwithstanding that the said judgment was premised unilaterally on one side.

It is certainly regrettable and unfortunate that the court in *Crinion* did not attempt to make any reference whatsoever to the Guide to Judicial Conduct in its decision. The Guide is the closest paraphernalia that judges have to a code of conduct without it actually being one.[47] It is not every day that an opportunity presents itself to the court which begets adjudication on judicial ethics. When the court does get such opportunity, one can almost expect the main non-jurisprudential source on the topic of judicial ethics in the United Kingdom would be cited. Hence, the omission in *Crinion* is patently disappointing because reference to the Guide would have prompted the court to further explicate the ethical principles at play and perhaps even encourage future courts to steer clear of certain prose and terminology.[48]

By contrast, Pauffley J of the Family Court in *Re L (A Child)* drew explicit attention to two out of the three principles ventilated in the Guide – independence and impartiality.[49] "It is difficult to view the justices as having been independent and impartial if, as happened here, [the court] simply adopted the local authority's analysis of what their findings and reasons might comprise."[50] Although Her Ladyship made no mention of the Guide *per se*, the relevant passages are nevertheless *in pari materia* with those principles enumerated in the non-

44 ibid, per Underhill LJ at [16] and Sir Stephen Sedley at [40].

45 ibid, at [17].

46 ibid, at [36].

47 ibid 7, at 4. The opening remarks sets out the purpose of the Guide, that is "to offer assistance to judges, coroners and magistrates about their conduct. It is based on the principle that responsibility for deciding whether or not a particular activity or course of conduct is appropriate rests with each individual judge." The remarks further stipulate that the Guide is "not a code, nor does it contain rules other than where stated. Instead, it contains a set of core principles which will help judges reach their own decisions."

48 Nothing more than a general remark was made by Sir Stephen Sedley, who stated at para 46: "I hope that a judgment like the one now before us will not be encountered again."

49 [2014] 1 WLR 2795.

50 ibid, at [68].

jurisprudential text. Furthermore, Her Ladyship thought that "it is fundamental that nothing is sent to the judge by one party unless it is copied simultaneously to every other party" in order to secure fairness to the parties.[51] A quick glimpse at the Guide reveals that exercising equality and fairness of treatment are part of the ethical principle of integrity.[52]

Turning back to *Crinion*, the message that the Court of Appeal is sending to judges is something along the lines of: "lifted judgments will be tolerated so long as you have properly addressed the case, the issues and the evidence bearing on the losing party.[53] Avoid extensive plagiarism though, as recriminations of bias and misconduct may arise more readily." Ultimately, the line that demarcates acceptable copying from inexcusable copying is extremely opaque.[54] There seems to be a tacit acceptance that judges may copy when counsel's submissions are of such quality that rewriting the reasoning and conclusions in the judge's own words would be such a waste of time.[55] In deciding as it did, the court in *Crinion* essentially preferred a lackadaisical approach to intellectual honesty – that judges, when delivering their judgments, are permitted to "fill up the empty vessel" first before deciding whether to engage in an elaborated disquisition of empirical analysis.[56]

Intellectual honesty, along with coherence and critical rigour, is a normative heritage of judicial ethics and discipline.[57] It is on this point that the court in *Crinion* left much to be desired. While lifted judgments may convey the extent of confidence that a judge holds in counsel's submissions,[58] this argument is fundamentally flawed and untenable because its inquiry is too limited. Suppose a judge is neither partial towards the winner or biased against the losing party, but instead lacked the requisite sophistication or conscientiousness to fully comprehend a particularly complex and protracted dispute. The matters arising from the dispute have never been adjudicated before and there are no established precedents. After hearing submissions from both sides, the judge delivers a judgment that reproduces a significant portion of counsel's submissions, making only inconsequential changes that afford little to no insight into the judge's own

51 ibid, at [67].

52 ibid 8, at 7.

53 *Crinion* (n 5).

54 See *English v Royal Mail Group Ltd* (2008), UKEAT/0027/08 where a verbatim reproduction of the respondents' submissions that completely ignored the appellant's submissions rendered the judgment of the Employment Tribunal fatal.

55 *Crinion* (n 5). For example, in para 5, Underhill LJ described the submissions of the counsel for the winning party as "thorough and carefully structured" and commended those submissions as "an excellent piece of work".

56 ibid. At para 16, Underhill LJ admitted that: "a judge will often derive great assistance from counsel's written submissions, and there is nothing inherently wrong in making extensive use of them, with proper acknowledgement, whether in setting out the facts or in analysing issues or the applicable legal principles or indeed in the actual dispositive reasoning."

57 ibid 25. In his article, Shapiro argues that all cooperative undertakings would be difficult or impossible in the absence of truthfulness.

58 ibid 55.

reasoning process. Surely, the lack of competence that the judge had ostensibly demonstrated cannot be said to be an exemplar of intellectual honesty. The courts frequently peddle the notion that justice must not only be done but must be seen to be done,[59] yet it is difficult to see how justice can be seen to be done in the scenario envisioned in light of the minimalist approach laid down in *Crinion*.[60]

Lurking beneath the rationale for minimalism is perhaps the apprehension of imposing more burden on judges who have very little control over their workload and that limited judicial resources would be further strained by meritless appeals based on make-weight allegations.[61] However, this burden must not be overstated for judicial accountability and the ethics of judgment writing seek "basic fairness, not perfection, and does not justify an undue shift in focus from the correctness of the result to an esoteric dissection of the words used to express the reasoning process behind it."[62] The pressure upon modern judges at both first instance and on appeal cannot be said to be greater than that of their forebears, even more so if one considers that judges of today are all accorded with the latest research apparatuses. An ethical judgment – one that encompasses independence, impartiality and integrity – need not be a lengthy judgment.[63] In fact, brevity is key to an authoritative and trenchant legal reasoning,[64] and at the same time allows judges to dispose their cases promptly.

In the end, practical realities support the conclusion that judicial plagiarism and ethical judgment writing simply cannot coexist. Whether the courts attract public support or criticism hinges on the quality of their reasons. The judicial duty to provide reasons can only be said to have been genuinely discharged if the reasons given truly reflect the views of the judge. Where the duty is largely circumvented as was the case in *Crinion*, the inequity that entails will prove difficult to be righted. In consonance with the right to fair trial[65] where decisions on litigated cases are neither submitted to nor blessed at the ballot box,[66] a plagiarised judgment bears the hallmark of a poisoned judgment, and a judge that projects such obvious moral turpitude inevitably drags the reputation of the bench into declension.

59 *R v Sussex Justices Ex p. McCarthy* [1924] 1 KB 256, 259. See also, *Cape Intermediate Holdings Ltd v Dring* [2020] AC 629; *Bank Mellat v HM Treasury* [2014] AC 700; *R v Abdroikov* [2007] 1 WLR 2679; *Porter v Magill* [2002] 2 AC 357; *R v Bow Street Metropolitan Stipendiary Magistrate Ex p. Pinochet Ugarte (No.2)* [2000] 1 AC 119.

60 Like Megarry once said: "To be condemned without being understood is as bad as being condemned unheard." See Robert Megarry, *Lawyer and Litigant in England* (Stevens & Sons 1962) 135.

61 ibid 33, at 390.

62 *R v Sheppard* [2002] 1 SCR 869, at [60].

63 Ward LJ was instructive on this point in *Baird v Thurrock Borough Council* [2005] EWCA Civ 1499 and opined: "Short judgments are, of course, all fine and well and to be encouraged but only if they are careful judgments."

64 ibid 8, at 215.

65 *English v Emery* (n 28) at [19]; *Anya v University of Oxford & another* [2001] IRLR 377 at [12]. For a broader overview of the jurisprudence of Article 6 of the European Convention of Human Rights, see *Garcia Ruiz v Spain* (2001) 31 EHRR 589; *Helle v Finland* (1997) 26 EHRR 159.

66 *Sheppard* (n 62), at [5].

IV. AN ALTERNATIVE TO THE MINIMALIST PROSE

Judgment writing is said to be "public writing of the highest order".[67] The question that must be asked is what do we expect of a judge? An appropriate response would be that ethical judgment writing intertwines style and substance,[68] and it is impossible to prescribe a formula of rigid methodology for crafting the perfect judgment.[69] While our expectations on the depth and precision of the judge's independent analysis must be guided by pragmatism over quixotism, it would not be unreasonable to demand that the judge's own imprimatur on the law,[70] at the most rudimentary level, must have explored both sides to a dispute and be capable of explaining to its audience where justice lies.[71]

For clarity, this paper neither attempts to endorse nor extol the idea of judicial originality. The underlying principle of *stare decisis* makes it impractical and undesirable to impose an originality requirement on the enterprise of judgment writing.[72] As one author puts it – "it is only the arrogant fool or the truly gifted who will depart entirely from the established template and reformulate an existing idea in the belief that in doing so they will improve it."[73] Drawing on Canadian jurisprudence, what is required instead is a "functional mechanism" that can determine whether the alleged deficiencies in reasons that a judgment contains effectively deprive a party of meaningful appellate review.[74] If the conclusion is in the affirmative, it follows that an error of law has been committed which warrants appellate intervention, and *vice versa*.[75]

The starting point for consideration would take into account a list of comprehensive factors including (1) the complexity of the dispute; (2) did the judge fully understand the intricacies of the dispute; (3) did the judge derive any assistance in drafting his or her findings;[76] (4) the extent of the judge's copying and what was copied;[77] (5) any significant inconsistencies or conflicts in evidence

67 ibid 14, at 237.
68 ibid, at 238.
69 ibid.
70 ibid, at 249.
71 ibid, at 309.
72 Co Litt 97b.
73 Duncan Webb, 'Plagiarism: A Threat to Lawyers' Integrity?' (*International Bar Association*, 2009) <http://www.ibanet.org/Article/Detail.aspx?ArticleUid=bc2ef7cd-3207-43d6-9e87-16c3bc2be595> accessed 26 June 2021.
74 *Sheppard* (n 62), at [25]
75 ibid.
76 *Virdi v Law Society (Solicitors Disciplinary Tribunal intervening)* [2010] 3 All ER 653. The Court of Appeal found that assistance rendered by the clerk to the tribunal in drafting their written findings was not *ultra vires* as the tribunal had no power to reconsider their decision.
77 *Re S* (n 36). 20% of the material in the judgment of first instance court which was taken verbatim or near verbatim from the skeleton arguments and written materials submitted by the parties was not serious enough to be appealed.

which are not addressed in the judgment;[78] (6) did the judge devote sufficient attention to the arguments of the complainant;[79] (7) whether the judge failed to take into account any material consideration or gave consideration to any immaterial circumstance;[80] (8) whether the judge clearly explained his preference for one case over the other;[81] (9) whether the copied text is supported by appropriate citations or up-to-date legal authority;[82] (10) did the judge deliberate or distinguish competing cases cited;[83] (11) any informal arrangement that might exist between the court and a litigant;[84] (12) any other intrinsic or extrinsic factor relevant to the determination of the exercise of independent analysis by the judge.

The multifactorial approach suggested above gives judges some leeway in preparing their judgments in that it is not a fine-tooth comb that sets an extremely lofty threshold of writing that reads as a work of art in itself. At the same time, the clemency granted to judges is not too lenient as to enable them to abdicate their core responsibility and to delegate the burden and cost of judgment writing to the parties. Where plagiarism is alleged, be it an unattributed inclusion of one paragraph or ten paragraphs dissipated sporadically throughout a 50-page long judgment or at a rate slightly below the 94% similarity level condemned in *Crinion*,[85] not all factors will be material and the weight assigned to the relevant factors may vary according to the facts of the dispute.

Returning to *Crinion*, the prospect of the impugned judgment being set aside is highly plausible had it been appraised against the list of factors detailed above. Of the 14 issues disputed, the trial judge did either one of these – made zero reference to the arguments ventilated by the defendant's counsel,[86] gave no reason as to why those arguments were rejected,[87] or substantially lifted passages from the claimant's submissions with extreme paucity of his own reasoning.[88] More egregiously, the "properties" segment of the electronic copy of the judgment readily revealed the author as counsel for the claimant.[89] Where the Court of Appeal was willing to overlook this mischief and to accept the minimalist prose of the first instance judge, this would not be palatable under the functional approach.

78 *Sheppard* (n 62) at [28].

79 *English* (n 5). A verbatim reproduction of the respondents' submissions that completely ignored the appellant's submissions rendered the judgment of the Employment Tribunal fatal.

80 *Re B* (n 10) at 347.

81 *Flannery* (n 28) at 382.

82 *Cojocaru v British Columbia Women's Hospital and Health Centre* [2013] 2 SCR 357 at [36]. The Canadian Supreme Court agreed with the view that a failure to attribute outside sources should be discouraged.

83 *Crinion* (n 5) at [17].

84 *Re L* (n 49). Pauffley J observed that in order to secure fairness to the parties and the perception that justice will be done, it is fundamental that nothing is sent to the judge by one party unless it is also circulated to the other party.

85 *Crinion* (n 5), at [11].

86 See generally, *IG Markets Ltd v Crinion* [2012] EWHC B4 (Mercantile).

87 ibid.

88 *Crinion* (n 5).

89 *Crinion* (n 5), at [11].

The ten cardinal factors call for a contextual and holistic consideration of all the circumstances which may have a bearing on the suggestion that the judge had indeed copied a partisan submission blindly and whether a fair-minded and informed observer would conclude that the judge had effectively abdicated his or her responsibility as a result of the copying.

In addition to the functional mechanism, this paper proposes that the following paragraph be inserted into the Guiding Principles of the Guide to Judicial Conduct that forms the wider notion of integrity:

> Judges are the official bearers of public trust and confidence in the courts. Therefore, the judgments that they write are held to high ethical standards. Judges must undertake intensive finding of fact and conclusion of law before arriving at a decision. A judge must not engage in extensive copying of partisan submissions and must ensure that no important evidence or argument from the other side is overlooked. Where a judge decides to borrow language from sources other than his own, the judge must do it in a way that does not foreclose a party of meaningful appellate review and must ensure that proper attribution is given. A judgment that fails to acknowledge borrowed language is a judgment lacking in integrity and reflects adversely on the ethics of the judiciary.

The inclusion of this proposed paragraph is not expected to be a silver bullet to every instance of judicial plagiarism, but it will provide a much-needed clarification to judges on the ethical boundaries of "cut-and-paste" judgments. It is not unrealistic to anticipate that a comment addressing plagiarism in the Guide will serve as a salutary deterrent against chameleon writing that adopts the winning litigant's prose and exhibits no distinctive thought or reasoning from the judge.[90] In doing so, this paragraph could pave the way for broaching the subject of lifted language that is often downplayed or goes unnoticed along the corridors of justice.

V. CONCLUSION

Judges are not rubber stamps that assent to the work of another as a substitute for their own. A superficial observation of the judicial process under the pretence of discharging judicial responsibility does not live up to the ethics and virtues envisioned in the Guide to Judicial Conduct. Construed narrowly, one side of a dispute which has not been given the closest personal attention by the judge

90 ibid 14, at 249.

renders the judicial process perfunctory.[91] In a cosmos where plagiarism is portrayed as *malum in se*, judges as the guardians of the rule of law are certainly not impervious to the stigma of disregarding this social imperative.

[91] ibid 2, at 211.

Cambridge Law Review (2022) Vol VII, Issue 2, 147–152

Assumptions of Irresponsibility: Liability for Omissions following *Tindall v Chief Constable of Thames Valley*

SAM PEARCE*

ABSTRACT

This case commentary analyses the present state of negligence liability in English tort law as set out in the recent case of *Tindall v Chief Constable of Thames Valley*.[1] Despite recent landmark decisions regarding acts and omissions, the boundaries of the distinction between the two remain to be fully explored. Following the decision in *Tindall*, it is suggested that a temporary conferral of a benefit must always fall to be classified as an omission. It is then argued that, for a claimant to establish that a defendant has assumed a responsibility to them, first it must be shown that the defendant has a relationship with the claimant that is sufficiently distinguishable from the general public. It is the lack of such a relationship that prevented the claimant in *Tindall* from successfully arguing that the police had assumed a responsibility to all road users. This commentary concludes that *Tindall* further elucidates key duty of care principles under the law of negligence, whilst also highlighting important questions that will require clarification from the courts in the future.

Keywords: negligence; duty of care; omissions; assumption of responsibility; public authorities

I. FACTS

After *Robinson v Chief Constable of West Yorkshire*[2], the position on when a public authority will owe a duty of care to an individual is no longer in flux. Settling a long line of conflicting case law, Lord Reed held, at [32], that "at common law,

* Student in BA (Hons) Law, University of Cambridge. I am grateful to the anonymous reviewers for their comments on earlier drafts. Any remaining errors are my own.

1 *Tindall v Chief Constable of Thames Valley* [2022] EWCA Civ 25 [2022] PIQR P10.

2 *Robinson v Chief Constable of West Yorkshire* [2018] UKSC 4, [2018] AC 736.

public authorities are generally subject to the same liability in tort as private individuals and bodies". This case note seeks to analyse the expansion of Lord Reed's position offered by the Court of Appeal concerning liability for omissions in *Tindall*.

The facts of *Tindall* are as follows: K, a driver, skidded on a patch of black ice and suffered non-life-threatening injuries. While waiting for the emergency services, K began to warn fellow road users about the dangerous, icy stretch of road. When the police arrived, K stopped warning other drivers and was taken to the hospital by ambulance. Meanwhile, the police erected a "Police Slow" sign and cleared the road of debris. Upon finishing, they retrieved the sign and exited the scene, leaving the black ice behind with no police presence to warn of its existence. Just twenty minutes later, T, another driver, collided with an individual who had lost control of their car on the ice. T died in the accident, and his widow brought a claim on his behalf

II. Decision and Comment

The Chief Constable appealed against the Master's refusal to strike out the claim at first instance, and succeeded in the Court of Appeal. Stuart-Smith LJ gave the leading judgment of the court, taking the opportunity to address the principles governing omissions liability comprehensively.

The initial distinction drawn, and one which must be drawn in any negligence claim, was whether the defendant's conduct amounted to an act or an omission.[3] Conduct that makes matters worse (at least, worse than if the defendant had done nothing), is generally considered an act. Such was the case in *Robinson* itself, when a group of policemen knocked into a frail and elderly woman whilst attempting to arrest a suspected drug dealer in a busy street. Conversely, omissions involve a failure to confer a benefit or a failure to prevent harm. The leading case is *Michael v Chief Constable of South Wales*[4], where a victim's emergency call to the police was given a lower priority than it should have had, resulting in the police's late arrival and their consequent discovery that the victim had already been stabbed to death by her former partner. In *Tindall*, Stuart-Smith LJ held that the police's conduct fell into the omissions category, dismissing two submissions made by the claimant in the process.

The first submission (at [67]) was that the police *had* made matters worse through their transient intervention of placing and then removing the "Police Slow" sign. Upon placing the sign, the police improved the situation by warning

[3] ibid [69(4)] (Lord Reed SCJ) "[A]lthough the distinction, like any other distinction, can be difficult to draw in borderline cases, it is of fundamental importance."

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

all road users about the dangerous condition of the road. When they subsequently removed the sign, the police made matters worse than they had been during the temporary period in which the sign was placed. The rationale for dismissing this submission is rooted in the idea that the defendant's removal of a temporary benefit they provided is not to be considered, in law, as a material worsening of the situation.[5] Stuart-Smith LJ's observation succinctly articulated the point, stating that "[the police] did not make matters worse: they merely left the road as they found it."[6] This observation does, however, invite further discussion. In avoiding confusion in more complex cases, it is worth discerning the limits of the concept of leaving something "as they found it", and if doing so should *always* be classified as failure to confer a benefit.

To illustrate the point, consider a situation where the police had, instead of erecting and removing the warning sign on an *ad hoc* basis as they did in *Tindall*, placed the sign down years earlier when dealing with another accident. Upon arriving at the scene and clearing debris off the road at present, would it then have been open to the police to retrieve the sign that they had placed so many years before? Two alternative answers appear available in response:

1. Leaving a situation "as they found it" is limited by temporal proximity to the improvement. After some arbitrary time period, a temporary intervention evolves into a permanent one, and its subsequent removal by the authority constitutes an act because it involves a worsening of the new state of affairs; or
2. Regardless of the elapsed time period, a defendant removing a benefit that they provided is always a failure to confer a benefit and must be construed as an omission.

Whilst the second option is demonstrably less generous towards claimants, it supports the general trend of case law that points away from finding liability when the actions of the defendant do not render individuals worse off than if the defendant had done nothing at all. It is submitted, therefore, that the second option reflects the current position of the law and, as Stuart-Smith LJ emphasised, no amount of incompetence on the part of the defendant in failing to confer the benefit or in removing the benefit that they provided can convert an omission into an act.

The second submission (at [66]) concerned whether the mere arrival of the police at the scene could give rise to a private law duty owed to road users to prevent them from harm. The claimant argued that, in coming to the accident,

5 *Capital and Counties Plc v Hampshire CC* [1997] QB 1004 (CA).
6 *Tindall v Chief Constable of Thames Valley* [2022] EWCA Civ 25, [2022] PIQR P10 [67] (Stuart-Smith LJ).

the police influenced K to leave in an ambulance, thereby causing him to cease providing warnings about the icy road to other drivers. Dispensing with the submission swiftly, Stuart-Smith LJ held that the police's contribution to K's decision to leave the scene by their mere arrival could not reasonably be described as negligent. McBride and Bagshaw have previously argued that certain case law, in contrast, supports the notion that a defendant who dissuades a third party from assisting a claimant can be held tortiously liable.[7] In *Costello v Chief Constable of Northumbria*[8], a police officer, B, was attacked and injured by a prisoner whilst a police inspector stood by and did nothing. A third officer in the station, H, may have been able to prevent the attack, but had left after noticing the inspector's presence and presuming that the inspector would step in should any violence break out. The authors suggest that the inspector's indication to H was a significant enough *interference* to give rise to a duty of care to B. By analogy, it is not too far a reach to suggest that, if the police had arrived and *told* K to stop warning drivers, they may have been interfering in a way that *was* negligent. As the police had given no such indication, and K had instead personally assumed that there was no longer any need for his presence at the scene, the claimant's argument was dismissed. The distinction is evidently a fine one, but it will be simply a matter for the court to determine whether or not a defendant's actions are meaningful enough to be considered an interference—and subsequently whether that interference can be described as negligent. Having rejected these submissions, the police's actions fell to be classified as an omission. The grounding of a duty of care in omissions cases relies upon the existence of a set of special circumstances beyond the presence of reasonable foreseeability of harm. Tofaris and Steel have summarised these circumstances as follows:

> In the tort of negligence, a person A is not under a duty to take care to prevent harm occurring to person B through a source of danger not created by A unless (i) A has assumed a responsibility to protect B from that danger, (ii) A has done something which prevents another from protecting B from that danger, (iii) A has a special level of control over that source of danger, or (iv) A's status creates an obligation to protect B from that danger.[9]

In *Tindall*, the claimant sought to rely on the first proposition as grounding a duty of care, namely that the police had assumed a responsibility to users on the road or to T himself by taking control of the scene and ineffectually handling the

[7] McBride and Bagshaw, *Tort Law* (6th edn, Pearson 2018) ch 6.4.

[8] *Costello v Chief Constable of Northumbria* [1999] 1 All ER 550.

[9] Tofaris and Steel, 'Negligence Liability for Omissions and the Police' (2016) 75 CLJ 128.

dangerous situation. Upholding this proposition, however, would have required an adverse manipulation of the concept of an "assumption of responsibility". An assumption of responsibility, as a legal term of art, is limited to specific situations such as when a contractual duty exists between the claimant and defendant or when a relationship akin to contract exists under the principle in *Hedley Byrne & Co Ltd v Heller & Partners Ltd*.[10] Absent of any features differentiating the relationship with the claimant from their relationship with anyone else, Stuart-Smith LJ affirmed the approach that, no matter how irresponsible the behaviour of a public authority is, they can never be said to have assumed a responsibility to the claimant.

This approach lends itself to a re-affirmation of the position established in *Kent v Griffiths*, a case where the ambulance service was held liable for failing to arrive on time to provide care for a patient suffering from bronchial asthma.[11] The decision in *Kent v Griffiths* has not been directly opposed by authority, but the precise ratio is worth discerning considering how, in principle, the ambulance service was held liable for failing to confer a benefit despite not assuming responsibility to the claimant personally. Applying the dicta of Stuart-Smith LJ, it is apparent that the duty of care in *Kent v Griffiths* is grounded by the emergency call to the ambulance service, which in turn establishes a relationship between the ambulance service and patient that is readily distinguishable from their relationship with the public at large.[12] There is no requirement, therefore, for a personal assumption of responsibility to the claimant, as all that is required is a sufficiently distinguishable relationship that the law can recognise as giving rise to a duty to act or confer a benefit. The features required to establish such a relationship will understandably differ under the circumstances of each case. It is safe to presume, though, that the bar for the weakest enforceable relationship requires at least bare knowledge of the existence of the person that the defendant would be assuming responsibility to.[13] In *Tindall*, the knowledge that road users, in general, would approach the dangerous patch of ice was insufficient to surpass that bar.

[10] *Stansbie v Troman* [1948] 2 KB 48 (CA); *Hedley Byrne & Co Ltd v Heller & Partners Ltd* [1964] AC 465 (HL). Liability under the *Hedley Byrne* principle has been confirmed to extend to omissions: *Midland Bank Trust Co Ltd v Hett Stubbs & Kemp* [1979] Ch 384 (Ch); *Henderson v Merrett Syndicates Ltd* [1995] 2 AC 145 (HL).

[11] *Kent v Griffiths* [2001] QB 36 (CA).

[12] For an alternative explanation premised on the concept of *interference* as discussed above, see McBride, 'Negligence Liability for Omissions - Some Fundamental Distinctions' [2006] Cambridge Student Law Review 10, 13.

[13] *Playboy Club London Ltd v Banca Nazionale del Lavoro SpA* [2018] UKSC 43, [2018] 1 WLR 4041. The UKSC held that there had been no assumption of responsibility under the *Hedley Byrne* principle to the claimant as the defendant had negligently supplied a favourable credit reference to the agent's undisclosed principal rather than to the claimant.

III. Conclusion

Stuart-Smith LJ's judgment demonstrates the clarity provided to personal injury claims following Lord Reed's dicta in *Robinson*. As Tofaris has indicated, *Robinson* provided the blueprint for the future development of the law of negligence, and *Tindall* is a decision that carefully places an additional building block upon that new blueprint.[14] Undoubtedly, there will be future cases where the acts-omissions distinction and the boundaries of the assumption of responsibility principle are more difficult to draw than they were here, but the helpful guidance established in *Tindall* will assist the courts in continuing to carve a more consistent path when faced with those challenging cases.

14 Tofaris, 'Duty of Care in Negligence: A Return to Orthodoxy?' (2018) 77 CLJ 454.

Cambridge Law Review (2022) Vol VII, Issue 2, 153–167

Dichotomy between Jurisdiction and Admissibility: Illuminating the Twilight Zone *BTN v BTP* [2021] 1 SLR 276

JOEL SOON*

ABSTRACT

The Singapore Court of Appeal's decision in *BTN v BTP* is significant insofar as it affirmed that the tribunal versus claim test, which was introduced in its earlier decision in *BBA v BAZ*, continues to apply to determine whether issues go towards jurisdiction or admissibility. Notwithstanding the strong impetus for drawing a dichotomy between jurisdiction and admissibility, the dichotomy's usefulness is called into question where issues defy easy classification. The inflexibility perpetuated by the dichotomy has led to the emergence of a twilight zone. This note will suggest that the dichotomy may be of limited usefulness in certain areas in the law of arbitration, but ultimately acknowledges that the Singapore courts are stuck between a rock and a hard place since alternatives have their own shortcomings.

Keywords: arbitration; jurisdiction; admissibility; twilight zone; Singapore

I. INTRODUCTION

After two seminal apex court decisions in *BBA v BAZ*[1] ('*BBA*') and *BTN v BTP*[2] ('*BTN*'), it is well-established in Singapore law that the 'tribunal versus claim' test, which asks whether the objection is targeted at the tribunal or the claim,[3] applies

* LLB Singapore Management University. I am grateful to the views and assistance of Ms Chang Wen Yee and Mr Louis Lau Yi Hang. All errors remain my own.

1 *BBA v BAZ* [2020] 2 SLR 453.

2 *BTN v BTP* [2021] 1 SLR 276.

3 *BBA* (n 1) [77]; *BTN* (n 2) [69].

to classify whether an issue goes towards jurisdiction or admissibility.[4] These decisions are to be welcomed for clarifying Singapore's approach to the dichotomy between jurisdiction and admissibility, which many have considered to be a 'longstanding issue' in international arbitration[5] where much ink has been spilled.[6]

While the dichotomy between jurisdiction and admissibility has been readily accepted by the Singapore courts, commentators have acknowledged that it is not always easy to establish a dividing line between jurisdiction and admissibility.[7] Indeed, there are cases where the dichotomy may be blurred,[8] making it difficult to fit the issue under either label.[9] In such cases where characterisation is not as straightforward, they are said to fall in a 'twilight zone'.[10] In this connection, eminent arbitration scholars, such as Hwang, have criticised the dichotomy between jurisdiction and admissibility, arguing to discard the 'admissibility' label in the 'tribunal versus claim' test.[11] The impetus for Hwang's argument stems from the failure of the test in elucidating *how* to identify if an objection targets the tribunal or the claim.[12]

In this note, the author scrutinises whether the dichotomy between jurisdiction and admissibility is useful by considering its application in several areas, ultimately concluding that trying to fit issues within either label may be redundant and akin to fitting a square peg into a round hole.[13] Attempts to do so will occasionally create an unnecessary twilight zone.

4 *BBA* (n 1) [76]; *BTN* (n 2) [69]. See also Margeret Joan Ling and Serene Chee, 'Recent Developments in Singapore Arbitration Law' (International Bar Association, 14 April 2021) <https://www.ibanet.org/article/F940CF84-99C9-4952-9BB1-94C24D5A42B9> accessed 22 November 2021.

5 Fabio G. Santacroce, 'Navigating the Troubled Waters Between Jurisdiction and Admissibility: An Analysis of Which Law Should Govern Characterization of Preliminary Issues in International Arbitration' (2017) 33(4) Arbitration International 539, 539.

6 Michael Hwang SC and Lim Si Cheng, 'The Chimera of Admissibility in International Arbitration – and Why We Need to Stop Chasing it' in *Selected Essays on Dispute Resolution* (SIAC Publishing, 2018) 431–475; Jan Paulsson, 'Jurisdiction and Admissibility' in Gerald Aksen *et al* (eds), *Global Reflections on International Law, Commerce and Dispute Resolution* (ICC Publishing, 2005) 608.

7 Paulsson, 'Jurisdiction and Admissibility' (n 6) 603, citing *Methanex Corporation v United States of America,* Partial Award on Jurisdiction and Admissibility, 7 August 2002, 7 ICSID Reports 239, 271; Andrew Tweeddale, 'Jurisdiction and Admissibility in Dispute Resolution Clauses' (2021) 16(1) Construction Law International 13, 14.

8 Yas Banifatemi, 'Chapter 1: The Impact of Corruption on 'Gateway Issues' of Arbitrability, Jurisdiction, Admissibility and Procedural Issues' in Domitille Baizeau and Richard Kreindler (eds), *Addressing Issues of Corruption in Commercial and Investment Arbitration* (ICC, 2015) 16, 19.

9 Santacroce, 'Navigating the Troubled Waters Between Jurisdiction and Admissibility' (n 5) 540; Tolu Obamuroh, 'Jurisdiction and Admissibility: A Case Study' (2020) 36(3) Arbitration International 373, 374.

10 Obamuroh, 'Jurisdiction and Admissibility' (n 9) 393–394; Paulsson, 'Jurisdiction and Admissibility' (n 6) 608; Luis Miguel Velarde Saffer and Jonathan Lim, 'Judicial Review of Investor Arbitration Awards: Proposals to Navigate the Twilight Zone between Jurisdiction and Admissibility' (2014) 8(1) Dispute Resolution International 85, 87 Santacroce, 'Navigating the Troubled Waters Between Jurisdiction and Admissibility' (n 5) 540.

11 Michael Hwang and Si Cheng Lim, 'Chapter 16: The Chimera of Admissibility in International Arbitration' in Neil Kaplan and Michael J. Moser (eds), *Jurisdiction, Admissibility and Choice of Law in International Arbitration* (Kluwer Law International, 2018) 265–288.

12 Hwang and Lim, 'The Chimera of Admissibility in International Arbitration – and Why We Need to Stop Chasing it' (n 6) 434.

13 Gretta Walters, 'Fitting a Square Peg into a Round Hole: Do Res Judicata Challenges in International Arbitration Constitute Jurisdictional or Admissibility Problems?' (2012) 29(6) Journal of International Arbitration 651.

II. Facts

A. Background

The first appellant, BTN, entered into a share purchase agreement with, *inter alios*, the respondents, BTP and BTQ, for the purchase of their interests in a group of companies. The share purchase agreement contained an arbitration clause stipulating the Singapore International Arbitration Centre's rules, and an exclusive jurisdiction clause for the Mauritian courts. It provided for the respondents' employment by the second appellant, BTO, under the Promoter Employment Agreements, which contained an arbitration clause also stipulating the Singapore International Arbitration Centre's rules, and an exclusive jurisdiction clause for the Malaysian courts. Under the share purchase agreement and Promoter Employment Agreements, the respondents could be terminated 'Without Cause' or 'With Cause'. Only the former entitles the respondents to a sum of money known as 'Earn Outs'.

B. Malaysian industrial court proceedings

Following the respondents' termination With Cause, proceedings were commenced before the Malaysian Industrial Court. After numerous adjournments of hearings owing to BTO's repeated absence,[14] the Malaysian Industrial Court found in favour of the respondents and ordered BTO to compensate them accordingly.[15] Despite some initial hesitance, BTO complied and effected full payment.[16]

C. Arbitration proceedings and decision below

The respondents then commenced arbitration under the share purchase agreement, claiming that their dismissal Without Cause entitled them to Earn Outs.[17] The main issue for the Tribunal's determination was the effect of the award rendered by the Malaysian Industrial Court.[18] This involved considering:[19] (a) what issues in the arbitration were said to be the subject of *res judicata*; and (b) whether the Malaysian Industrial Court's findings were binding on the Tribunal,

[14] *BTN* (n 2) [18].
[15] ibid [19].
[16] ibid [24].
[17] ibid [25].
[18] ibid [27].
[19] ibid [27].

in that the issues dealing 'with cause of termination' were *res judicata* because of the award rendered by the Malaysian Industrial Court.[20] The Tribunal, in its Partial Award, held that the issue estoppel doctrine under Singapore law prevented the appellants from arguing that the respondents were terminated 'With Cause', as this was effectively determined by the Malaysian Industrial Court.[21]

The appellants' application to the High Court was dismissed by the judge,[22] who held, *inter alia*, that the Partial Award was not a ruling on jurisdiction, as the *res judicata* issue was not a jurisdictional issue.[23] Additionally, the Partial Award was not contrary to Singapore's public policy, as the appellants had their case heard.[24]

D. The Court of Appeal's Decision

The appellants appealed and argued, *inter alia*, that the Partial Award was contrary to Singapore public policy for two reasons.[25] First, their ignorance of the Malaysian Industrial Court proceedings deprived them of the right to defend themselves and/or make claims relating to the respondents' termination 'With Cause' under the share purchase agreement.[26] Secondly, upholding the Partial Award would allow the respondents to take advantage of their purported breach of the Promoter Employment Agreements' arbitration agreement.[27]

The Court of Appeal rejected both arguments. It held that the appellants' alleged ignorance of the Malaysian Industrial Court proceedings was irrelevant because it resulted from its own internal arrangements.[28] They are precluded from refusing to accept the Tribunal's determination, or from complaining about the Tribunal's failure to conduct a factual inquiry into the circumstances behind BTO's non-appearance at the Malaysian Industrial Court proceedings – any relevant challenge or argument could have been made before the Tribunal.[29] The second argument was unmeritorious,[30] as the mandatory nature of the arbitration clause was conditional on one party invoking it. Short of this, the actions of the respondents taken in relation to the Malaysian Industrial Court did not breach the arbitration agreement.[31]

[20] ibid [26].
[21] ibid [33]–[34].
[22] ibid [35].
[23] ibid [36].
[24] ibid [36].
[25] ibid [37] and [57].
[26] ibid [57].
[27] ibid [57].
[28] ibid [59].
[29] ibid [59]–[61].
[30] ibid [63].
[31] ibid [63].

Pertinently, the Court of Appeal addressed the appellants' additional argument: if the award rests on an error of law (in this case, erroneous applications of *res judicata*) that resulted in the Tribunal not exercising its mandate, the award should be set aside on the public policy ground.[32] However, the Court of Appeal noted that 'errors of law or fact made in an arbitral decision, *per se*, are final and binding on the parties',[33] and do not engage Singapore public policy.[34] Conversely, a tribunal's decision on jurisdiction is subject to *de novo* independent review by the courts.[35]

In this connection, the appellants suggested that the present tribunal's decision that it was unable to exercise its mandate, was a decision on jurisdiction. The Court of Appeal disagreed. Applying the 'tribunal versus claim' test,[36] the Court of Appeal held that a tribunal's decision on the *res judicata* effect of a prior decision is not a decision on jurisdiction, but rather an issue on admissibility.[37] As explained in *The Royal Bank of Scotland NV v TT International Ltd*,[38] which laid down principles equally applicable to a tribunal's decision on *res judicata*,[39] the *res judicata* doctrine operates against litigants, and not against courts.[40] It does not have any effect on the court's authority to hear the dispute before it.[41] Further, where a party argues that a dispute has already been resolved, the party is not seeking resolution of that dispute in another forum; instead, the party does not want the claim to be resolved in any forum.[42] Accordingly, the appellants' jurisdictional challenge failed on this distinction, because *res judicata* issues go towards admissibility.[43]

III. Analysis

A. The dichotomy between jurisdiction and admissibility

Before examining how the Court of Appeal in *BTN* applied the 'tribunal versus claim' test, it is pertinent to explore the dichotomy between jurisdiction and

32 ibid [65].

33 ibid [66]; *PT Asuransi Jasa Indonesia (Persero) v Dexia Bank SA* [2007] 1 SLR(R) 597, [56]. Cf. *Oil & Natural Gas Corporation Ltd v Saw Pipes Ltd* AIR 2003 SC 2629.

34 *BTN* (n 2) [66]; *PT Asuransi* (n 33) [57]; *AJU v AJT* [2011] 4 SLR 739, [66].

35 *BTN* (n 2) [66].

36 ibid [69], citing *BBA* (n 1) [77]–[79].

37 *BTN* (n 2) [71].

38 *Royal Bank of Scotland NV v TT International Ltd* [2015] 5 SLR 1104.

39 *BTN* (n 2) [71].

40 *Royal* (n 38) [115].

41 ibid [115].

42 *BTN* (n 2) [71], citing Walters, 'Fitting a Square Peg into a Round Hole' (n 13) 675.

43 *BTN* (n 2) [71], and [74]–[77] where the Court of Appeal disagreed with reasoning from two foreign cases because they stand for a position which the Court of Appeal does not accept in Singapore.

admissibility in greater detail,[44] to understand the implications that flow therefrom. While jurisdiction refers to 'the power of the tribunal to hear a case', admissibility asks the question of 'whether it is appropriate for the tribunal to hear it'.[45] They are similar in no less than two ways: they are not only both part of the universe of preliminary questions,[46] but a finding of either lack of jurisdiction and/or inadmissibility will lead to the same result – the tribunal withholds itself from examining the merits of the claim.[47] Despite the similarities, the fundamental distinction between the two concepts is significant,[48] and is 'not merely an exercise in linguistic hygiene pursuant to a pedantic hair-splitting endeavour'.[49] As a tribunal's jurisdiction is founded on the parties' consent,[50] to object against an arbitral tribunal's jurisdiction is to argue that consent is non-existent, invalid, not within the scope of the dispute in issue, or in violation of public policy.[51] However, when an admissibility challenge is raised, the party alleges that a claim is defective, and should not be heard in any forum.[52] Examples include timeliness, mootness, and ripeness.[53] As there is inherent difficulty in determining whether an objection goes to jurisdiction or admissibility, the 'tribunal versus claim' test attempts to simplify this exercise: objections attacking the tribunal are classified as jurisdictional in nature, and those targeting the claim are objections to admissibility.[54]

Although the Court of Appeal in *BTN* applied the 'tribunal versus claim' test without much difficulty, the author submits that it may not always provide helpful guidance in distinguishing between an objection to jurisdiction or

[44] Nikita V Nota, 'International Arbitration: Some Reflections on Jurisdiction and Admissibility' (2010) 2 Ukrainian Journal of Business Law 31, 31; Santacroce, 'Navigating the Troubled Waters Between Jurisdiction and Admissibility' (n 5) 539.

[45] See Zachary Douglas, *The International Law of Investment Claims (Cambridge University Press, 2009) [291] and [310];* Chin Leng Lim, Jean Ho and Martin Paparinskis, *International Investment Law and Arbitration: Commentary, Awards and Other Materials* (Cambridge University Press, 2018), 118. See Obamuroh, 'Jurisdiction and Admissibility' (n 9) 377. Tweeddale, 'Jurisdiction and Admissibility in Dispute Resolution Clauses' (n 7) 13–14.

[46] Nota, 'International Arbitration' (n 44) 32.

[47] Hwang and Lim, 'The Chimera of Admissibility in International Arbitration – and Why We Need to Stop Chasing it' (n 6) 433; Walters, 'Fitting a Square Peg into a Round Hole' (n 13) 661; Santacroce, 'Navigating the Troubled Waters Between Jurisdiction and Admissibility' (n 5) 540; Yas Banifatemi, 'Chapter 1' (n 8) 19.

[48] Paulsson, 'Jurisdiction and Admissibility' (n 6) 603.

[49] *Swissbourgh Diamond Mines (Pty) Ltd and others v Kingdom of Lesotho* [2019] 1 SLR 263, [208].

[50] N Blackaby, C Partasides QC, A Redfern, and M Hunter, *Redfern and Hunter on International Arbitration* (6th edn, Oxford University Press 2015) [5.110]; Emmanuel Gaillard and John Savage, *Fouchard Gaillard Goldman on International Commercial Arbitration* (2nd edn, Kluwer Law International 1999), 253.

[51] *BBA* (n 1) [78]; Walters, 'Fitting a Square Peg into a Round Hole' (n 13) 661.

[52] Hanno Wehland, 'Jurisdiction and Admissibility in Proceedings Under the ICSID Convention and ICSID Additional Faculty Rules' in Crina Baltag (ed), *ICSID Convention After 50 Years: Unsettled Issues* (Kluwer Law International, 2016) 227, 234; Walters, 'Fitting a Square Peg into a Round Hole' (n 13) 661; Nota, 'International Arbitration' (n 44) 32.

[53] Obamuroh, 'Jurisdiction and Admissibility' (n 9) 391; William W Park, 'Determining an Arbitrator's Jurisdiction: Timing and Finality in American Law' (2007) 8 Nevada Law Journal 135, 153; Walters, 'Fitting a Square Peg into a Round Hole' (n 13) 662.

[54] Hwang and Lim, 'The Chimera of Admissibility in International Arbitration – and Why We Need to Stop Chasing it' (n 6) 433; Paulsson, 'Jurisdiction and Admissibility' (n 6) 616.

admissibility.[55] In arriving at the correct conclusion that *res judicata* issues are admissibility issues,[56] the Court of Appeal relied on principles laid down in *Royal Bank*, which involved *res judicata* in the context of court proceedings, and on well-reasoned 'logic' as explained by Gretta Walters.[57] While the Court of Appeal suggested that 'this statement of principle is applicable to decisions made by arbitral tribunals on issues of *res judicata*',[58] ultimately it did not directly apply the 'tribunal versus claim' test to explain how *res judicata* attacks the claim in the context of arbitration proceedings. One could, on this basis, question the efficacy of the test.

Since there exists no clear guidance in academic literature as to *when* an objection targets the claim or tribunal, it has been suggested that 'instincts' are possibly relied upon when making such determination.[59] However, courts should be wary of such unsatisfactory forms of decision-making, 'because it [would] involve a veiled reliance on instinct which is sheltered from scrutiny as opposed to express reasoning'.[60]

To be clear, the Court of Appeal most certainly averted such problems. It was also was fully entitled to rely on *Royal Bank* to reach its conclusion on the *res judicata* issue. But good judicial decision-making on one occasion does not necessarily cure the inadequacy of the 'tribunal versus claim' test. As will be discussed, where claims involve conditions precedent or non-arbitrability, it could be argued that they lie within the twilight zone where the answer is not crystal clear,[61] and application of the 'tribunal versus claim' test thereto may not yield the same success in terms of classification.

B. Areas where the Dichotomy between Jurisdiction and Admissibility may not be Useful

(i) Conditions Precedent to Arbitration

Multi-tier dispute resolution clauses,[62] which provide for arbitration only

55 Hwang and Lim, 'The Chimera of Admissibility in International Arbitration – and Why We Need to Stop Chasing it' (n 6) 434.

56 *BTN* (n 2) [71]. See also *Chiro Corp. v Ortho Diagnosis Sys.*, 207 F.3d 1126 (9th Cir. 2000); *Marriott International Hotels, Inc v J.N.A.H. Development S.A.* (2010) no. 09/13559.

57 *BTN* (n 2) [71], citing Walters, 'Fitting a Square Peg into a Round Hole' (n 13) 672 and 675.

58 *BTN* (n 2) [71].

59 Hwang and Lim, 'The Chimera of Admissibility in International Arbitration – and Why We Need to Stop Chasing it' (n 6) 454.

60 ibid 455.

61 Miguel and Lim, 'Judicial Review of Investor Arbitration Awards' (n 10) 89.

62 For clarity, such clauses are enforceable. See *International Research Corp. PLC v Lufthansa Systems Asia Pacific Pte Ltd* [2012] SGHC 226; *HSBC Institutional Trust Service v Toshin Development Singapore Pte Ltd* [2012] 4 SLR 378. See also *United Group Rail Services Ltd v Rail Corporation New South Wales* (2009) 127 Con LR 202; *Emirates Trading Agency LLC v Prime Minister Exports Pte Ltd* [2015] 1 WLR 1145. Cf. *Walford v Miles* [1992] 2 AC 128.

after contractually-prescribed procedures have been exhausted (conditions precedent to arbitration),[63] are increasingly being adopted, especially in complex construction and engineering contracts.[64] Despite such clauses being attractive for promoting efficiency, cost-savings and cooperation,[65] they are notoriously known as 'midnight clauses'[66] which are inserted at the eleventh-hour of contractual negotiations.[67] Unsurprisingly, multi-tier dispute resolution clauses tend to be haphazardly drafted.[68]

This is significant because the construction of such clauses can affect whether it is a jurisdictional or admissibility issue.[69] Whereas it could be regarded as 'jurisdictional' on the theory that it is a condition to a party's consent to arbitrate, it could also be characterised as an admissibility issue because the claim is not ripe to be heard.[70] Indeed, this is an area where national and international authorities have diverged.[71] Such dissonance reveals that the application of the dichotomy is not so straightforward.

The UK decision in *The Republic of Sierra Leone v SL Mining Ltd*[72] is an apt starting point. There, Sir Michael Burton noted that '[t]he views of the leading academic writers, [were] all one way',[73] as with 'important decisions in other jurisdictions'[74] – the failure to satisfy conditions precedent to arbitration, is a question of admissibility.[75] Thereafter, the Hong Kong court in *C v D*[76] followed *SL Mining*, reaching the same conclusion on the issue.[77] Notably, the court held that there was no dispute about the existence, scope and validity of the arbitration

63 Gary Born, *International Commercial Arbitration* (2nd edn, Kluwer Law International, 2014) 278.

64 Michael Pryles, 'Multi-Tiered Dispute Resolution Clauses' (2001) 18(2) Journal of International Arbitration 159, 159.

65 ibid.

66 Jo Delaney and Charlotte Hendriks, 'Multi-Tiered Dispute Resolution Clauses: A Reminder of the Court of Appeal's Split Decision' *Global Arbitration News* (11 August 2020) <https://globalarbitrationnews.com/multi-tiered-dispute-resolution-clauses-a-reminder-of-the-court-of-appeals-split-decision/> accessed 22 November 2021.

67 Didem Kayali, 'Enforceability of Multi-Tiered Dispute Resolution Clauses' (2010) 27(6) Journal of International Arbitration 551, 553.

68 ibid.

69 Michael McErlaine and James Allsop, 'Trends in Questions of Jurisdiction and Admissibility in International Arbitration' (Kluwer Arbitration Blog, 2 November 2021) <http://arbitrationblog.kluwerarbitration.com/2021/11/02/trends-in-questions-of-jurisdiction-and-admissibility-in-international-arbitration/> accessed 22 November 2021.

70 Gary Born, *International Commercial Arbitration* (3rd edn, Kluwer Law International 2021) 997–998.

71 ibid 999.

72 *The Republic of Sierra Leone v SL Mining Ltd* [2021] EWHC 286.

73 Born, *International Commercial Arbitration* (n 70) 1000; Jan Paulsson, 'Jurisdiction and Admissibility' (n 6) 616–617; Alex Mills, *Party Autonomy in Private International Law* (Cambridge University Press, 2018) [6.4.1]; Alex Mills, 'Arbitral Jurisdiction' in Thomas Schultz and Federico (eds), *Oxford Handbook of International Arbitration* (Oxford University Press, 2020) 6–7.

74 *BG Group v Republic of Argentina* 134 S.Ct.1198 (2002) (US Supreme Court); *BBA* (n 1); *BTN* (n 2).

75 *SL Mining* (n 72) [14]–[15] and [21]. Cf *Emirates* (n 62), where although the court suggested that such issues are a matter of jurisdiction, it was not cognisant of the Dichotomy, and hence should not be relied upon.

76 *C v D* [2021] HKCFI 1474.

77 ibid [42] and [53], citing *SL Mining* (n 72) [16].

agreement, and the parties' commitment to arbitrate was not in doubt.[78] Recently, Calver J in *NWA v NVF*[79] applied Sir Michael Burton's reasoning in *SL Mining*, finding that questions of whether a clause amounted to a condition precedent and whether it had been breached were matters of admissibility.[80] In the court's view, such an approach, as advocated in academic commentaries,[81] is consistent with both the commercial purpose of arbitration clauses[82] and the objective intention of the parties.[83]

However, the position is far from settled. The Singapore Court of Appeal in *International Research Corp PLC v Lufthansa Systems Asia Pacific Pte Ltd* ('***IRC***') found that the tribunal did not have 'jurisdiction' to proceed with the arbitration, due to non-compliance with the multi-tier dispute resolution clause.[84] Although *IRC* may seem to be at odds with the English and Hong Kong positions, this is likely because *IRC* was decided in a time when the dichotomy between jurisdiction and admissibility had not yet been adverted to in a Singapore case; such a consideration was also noted by Sir Michael Burton when he analysed English authorities preceding *SL Mining*.[85] Given the weight of English and Hong Kong authorities, the author posits that the Singapore courts would likely consider that non-compliance with multi-tier dispute resolution clauses fall within the 'admissibility' label,[86] notwithstanding its lack of opportunity to do so till date.

In any event, the lack of coherence in how multi-tier dispute resolution clauses have been characterised is noticeable.[87] Although it could plausibly be argued that the Swiss Supreme Court's ruling that arbitration proceedings should be stayed until pre-arbitral steps have been complied with[88] is a reference to admissibility, the Swiss courts have equivocated in this regard, given that they have

78 ibid [53]. See also Born, *International Commercial Arbitration* (n 70) 1007.

79 *NWA v NVF* [2021] EWHC 2666.

80 ibid [55] and [67]. Cf. *Emirates* (n 62) and *Tang v Grant Thornton International Limited* [2013] 1 All ER 1226.

81 *NWA* (n 79) [48]–[53], citing Born, *International Commercial Arbitration* (n 70) 975 and 1000; Louis Flannery QC and Robert M Merkin QC, *Merkin & Flannery on the Arbitration* (6th edn, Informa Law 2019) [30.13.2]; Chartered Institute of Arbitrators, International Arbitration Practice Guideline (2015) <https://www.ciarb.org/media/4192/guideline-3-jurisdictional-challenges-2015.pdf> accessed 22 November 2021; Paulsson, 'Jurisdiction and Admissibility' (n 6) 614–617.

82 *Premium Nafta Products Limited v Fili Shipping Company Limited* [2007] UKHL 40, [5]–[8].

83 *NWA* (n 79) [47].

84 *International Research Corp PLC v Lufthansa Systems Asia Pacific Pte Ltd* [2014] 1 SLR 130, [63]; Nandakumar Ponniya and Michelle Lee, 'Chapter 9: Commencement of Arbitration' in Sundaresh Menon (ed), *Arbitration in Singapore – A Practical Guide* (2nd edn, Sweet & Maxwell 2018) [9.022].

85 *SL Mining* (n 72) [13].

86 This is a position that is indeed consistent with UK and Hong Kong. Both *SL Mining* and *C v D* had cited Singapore authorities and applied the Test that was endorsed in Singapore in coming to its eventual conclusion. This arguably demonstrates how the Test would likely be applied by the Singapore courts if and when the time comes. See *SL Mining* (n 72) and *C v D* (n 76).

87 See generally, Hamish Lai *et al*, 'Multi-Tiered Dispute Resolution Clauses in International Arbitration – The Need for Coherence' (2020) 38(4) ASA Bulletin 796.

88 *X Ltd v Y SpA* [2016] 4A_628/2015, 18.

used the labels 'admissible' and 'jurisdictional' synonymously.[89] This conflation is unsurprising if one considers that the dichotomy between jurisdiction and admissibility has not played a major role in Swiss commercial arbitration. The focus has simply been on jurisdiction.[90] Similarly, it is unclear in Australia whether non-compliance with multi-tier dispute resolution clauses is a jurisdictional or admissibility issue.[91] Paulsson even suggests that the failure to respect condition precedents could be a jurisdictional issue, if a party insists that his consent to arbitration is contingent on a *bona fide* attempt at settlement.[92] With the characterisation of conditions precedent to arbitration varying among different legal systems,[93] perhaps owing to the difference in how civil law and common law lawyers look at this issue,[94] the dichotomy between jurisdiction and admissibility may be of limited usefulness in that it serves to obfuscate rather than explain. One may perhaps see the 'tribunal versus claim' test as a crude attempt to pigeonhole legal principles into either admissibility or jurisdiction, when they could shade into either depending on the perspective adopted or the facts of each case.

(ii) Non-arbitrability

Non-arbitrability is another area in which the dichotomy between jurisdiction and admissibility may be unhelpful. Generally, arbitrability refers to the possibility or otherwise of settling a dispute by arbitration.[95] Although what amounts to arbitrable subject matter is not the subject of comprehensive statutory guidance,[96] the 'concept of arbitrability finds legislative expression in section 11 of the IAA',[97] where subject matter arbitrability is subject to the limits imposed by public policy.[98] Singapore has thus chosen to define areas of non-arbitrability by

89 Tweeddale, 'Jurisdiction and Admissibility in Dispute Resolution Clauses' (n 7) 15, citing *A SA v B SA* [2014] 4A_124/2014, where the Swiss Supreme Court was also unclear as to whether non-compliance with FIDIC's mandatory requirements gave rise to a jurisdictional or admissibility challenge.

90 Marco Stacher, 'Jurisdiction and Admissibility under Swiss Arbitration Law – the Relevance of the Distinction and a New Hope' (2020) 38(1) ASA Bulletin 55, 55.

91 George M Vlavianos and Vasilis F L Pappas, 'Multi-Tier Dispute Resolution Clause as Jurisdictional Conditions Precedent to Arbitration' *Global Arbitration Review* (6 June 2017) <https://globalarbitrationreview.com/guide/the-guide-energy-arbitrations/2nd-edition/article/multi-tier-dispute-resolution-clauses-jurisdictional-conditions-precedent-arbitration> accessed 22 November 2021, citing *United Group Rail Services Ltd v Rail Corp New South Wales* [2009] NSWCA 177.

92 Paulsson, 'Jurisdiction and Admissibility' (n 6) 613.

93 Born, *International Commercial Arbitration* (n 70) 989.

94 Tweeddale, 'Jurisdiction and Admissibility in Dispute Resolution Clauses' (n 7) 14.

95 Hunter, *Redfern and Hunter on International Arbitration* (n 50) [2.29]; Obamuroh, 'Jurisdiction and Admissibility' (n 9) 386; Loukas A Mistelis, 'Arbitrability – International and Comparative Perspectives' in Loukas A Mistelis and Stavros Brekoulakis (eds), *Arbitrability: International and Comparative Perspectives* (Kluwer Law International, 2009) 1, 3–4.

96 Menon, *Arbitration in Singapore* (n 84) [15.056].

97 *Westbridge Ventures II Investment Holdings v Anupam Mittal* [2021] SGHC 244, [25].

98 ibid [25]. See also *Larsen Oil and Gas Pte Ltd v Petroprod Ltd* [2011] 3 SLR 414, [28]–[30].

with reference to Singapore public policy.[99]

Whether non-arbitrability is a jurisdictional or admissibility issue was most recently considered in *Westbridge Ventures II Investment Holdings v Anupam Mittal*,[100] where Mohan JC (as he then was) ruled that a finding of arbitrability (or non-arbitrability) is one that strikes at the tribunal's jurisdiction in respect of that dispute.[101] In reaching this conclusion, reliance was placed on several authorities which state that arbitrability is a question of jurisdiction. An excerpt from Bernard Hanotiau's article was cited, which stated that '[a]rbitrability is indeed a condition of validity of the arbitration agreement and, consequently, of the arbitrator's jurisdiction'.[102] Further, the High Court observed that this point was echoed in *Comparative International Commercial Arbitration*: '[t]hough arbitrability is often considered to be a requirement for the validity of the arbitration agreement it is primarily a question of jurisdiction'.[103] With respect, however, to the extent that these authorities had not explicitly considered the dichotomy, they may be of limited value in determining whether the issue of arbitrability goes towards jurisdiction or admissibility.

More crucially, the High Court had recourse to the 'tribunal versus claim' test,[104] and found that non-arbitrability raises a defect as to the parties' consent to arbitration.[105] Parties' consent would be invalid[106] since parties cannot agree to submit non-arbitrable disputes to arbitration as a matter of public policy.[107] The High Court observed that the issue of subject matter arbitrability 'cannot merely be a matter of admissibility'; instead, it strikes at the tribunal's jurisdiction.[108] While the assumption that matters which do not go to admissibility necessarily go to jurisdiction in this context, this does not preclude an interpretation that non-arbitrability could possibly be *both* a matter of admissibility and jurisdiction,[109] depending on the circumstances at hand.

First, Menon CJ, speaking extrajudicially, noted that the doctrine of non-arbitrability 'is not an indictment of the ability of arbitrators to deal with such

99 Menon, *Arbitration in Singapore* (n 84) [15.057]. It should be noted this notion of public policy is potentially broader than that identified in Article 34(2)(b)(ii) of the Model Law, which is concerned with fundamental notions and principles of justice as opposed to domestic policy considerations. See also *Westbridge* (n 97) [26]. Cf. *PT Asuransi Jasa Indonesia (Persero) v Dexia Bank SA* [2007] 1 SLR(R) 597, [59].

100 *Westbridge* (n 97).

101 ibid [36].

102 Bernard Hanotiau, 'The Law Applicable to Arbitrability' (2014) 26 SAcLJ 874, [1].

103 Julian David, Mathew Lew QC, Loukas A Mistelis and Stefan Kröll, *Comparative International Commercial Arbitration* (Kluwer Law International, 2003) [9-18].

104 *Westbridge* (n 97) [39]–[40].

105 ibid [40].

106 The arbitration agreement would be rendered 'inoperative' or 'incapable of being performed' under section 6 of the IAA. See also *Tomolugen Holdings Ltd v Silica Investors Ltd* [2016] 1 SLR 373, [73].

107 *Westbridge* (n 97) [40].

108 ibid [41]. See also Stacher, 'Jurisdiction and Admissibility under Swiss Arbitration Law' (n 90) 57, where arbitrability is also a jurisdictional issue in the Swiss courts.

109 Paulsson, 'Jurisdiction and Admissibility' (n 6) 614.

issues, but simply a reflection of the limits of arbitration rooted in contract'.[110] Insofar as jurisdiction refers to 'the power of the tribunal to hear a case', it could be argued that non-arbitrability may not fall neatly within the jurisdiction label. Secondly, Paulsson suggested that the US Supreme Court, in two cases,[111] implicitly treated the issue of arbitrability as an admissibility issue.[112] That there is no panacea can be observed from the fact that he goes on to question whether the classification was correct in these cases, suggesting that non-arbitrability could also go to the issue of jurisdiction.[113] For now, while Singapore jurisprudence has had the fortune of *Westbridge's* guidance on this issue, it can at least be said that the issue of non-arbitrability is one that defies easy classification, insofar as the dichotomy between jurisdiction and admissibility is concerned.

C. An alternative?

It is apposite to turn to consider the alternatives which have been proffered. According to Hwang, the question should simply be 'whether the objection, if factually proven, would impinge upon the *consent* of the objecting party to arbitration, so as to amount to a jurisdictional objection'.[114] Instead of force-fitting issues into a binary between jurisdiction and admissibility, Hwang suggests that the inquiry should be whether an objection is jurisdictional or non-jurisdictional.[115] This stands in stark contrast to the present tribunal versus claim test, which equates non-jurisdictional inquiries with admissibility when it may not necessarily be as straightforward.

As Hwang posits, adjudicators should 'open [their] minds to alternative methods which may be better at identifying if consent is affected.[116] On conditions precedent, the suggested approach is to apply contractual interpretation to '*interpret the offer to arbitrate*' to determine if the party had 'intended the precondition to be a *condition to its consent* to arbitrate' [emphasis in original].[117] Such an approach comports not only with the Singapore courts' focus on the

110 Sundaresh Menon, 'Arbitration's Blade: International Arbitration and the Rule of Law' (2021) 38(1) Journal of International Arbitration 1, 9.

111 *Howsam v Dean Witter Reynolds, Inc* (2002) 536 U.S. 79; *Green Tree Financial Corp. v Bazzle* (2003) 123 S Ct 2402.

112 Paulsson, 'Jurisdiction and Admissibility' (n 6) 612.

113 ibid 613.

114 Hwang and Lim, 'The Chimera of Admissibility in International Arbitration – and Why We Need to Stop Chasing it' (n 6) 434.

115 See also Stacher, 'Jurisdiction and Admissibility under Swiss Arbitration Law' (n 90), where the author proposes not to use the jurisdiction-admissibility dichotomy for Swiss law, but to focus on whether an issue is a jurisdictional one.

116 Hwang and Lim, 'The Chimera of Admissibility in International Arbitration – and Why We Need to Stop Chasing it' (n 6) 434–435.

117 ibid 435. See also Tweeddale, 'Jurisdiction and Admissibility in Dispute Resolution Clauses' (n 7) 15.

underlying 'substance'[118] of the 'tribunal versus claim' test where analysing parties' consent is paramount,[119] but also with the English courts' focus on giving effect to the commercial purpose of the arbitration clause[120] and the 'objective intention of the parties'.[121] Also in line with this is Born's suggestion that the consequences of non-compliance with conditions precedent 'are ultimately matters of contractual interpretation'.[122]

It is proposed that the 'tribunal versus claim' test can remain of general application, but where courts are faced with scenarios where the dichotomy between jurisdiction and admissibility is less clear, Hwang's approach would circumvent problems posed by the twilight zone. The upshot of this proposal is that it averts tying the adjudicators' hands into conclusively placing an issue within the binary, where the answer may not strictly lie therein. In practice, cases such as *BTN* and *BBA* will be unaffected as they can be resolved solely on an application of the 'tribunal versus claim' test. But where this proposed approach comes in handy is where issues defy easy classification under the dichotomy. For issues that fall outside jurisdiction but cannot be clearly said to be one of admissibility, Hwang's approach would label it as a non-jurisdictional issue, thereby leaving no room for the twilight zone.

Such an approach encourages principled decision-making. For jurisdictional issues, the body of rules concerning jurisdiction can continue to apply.[123] Conversely, for non-jurisdictional issues, the tribunal has 'the discretion to conduct proceedings in such manner as it considers appropriate'.[124] As for difficulties associated with Hwang's approach, it is worthy of another discussion in and of itself. Briefly, it is suggested that such an approach would lack certainty, since courts would invariably have to decide on *how* to identify if consent is affected.

D. Impetus for the dichotomy between jurisdiction and admissibility

Despite the potential difficulty in classifying certain issues under either label, as alluded to above, the author acknowledges that the dichotomy between jurisdiction and admissibility remains relevant for several reasons. First, it has been

118 *Westbridge* (n 97) [40].

119 *BBA* (n 1) [78].

120 *NWA* (n 79) [33], citing *Premium Nafta* (n 82) [5]–[8].

121 ibid [47].

122 ibid [48], citing Born, *International Commercial Arbitration* (n 70) 975.

123 Hwang and Lim, 'The Chimera of Admissibility in International Arbitration – and Why We Need to Stop Chasing it' (n 6) 435. See, *eg*, Rule 28 of the Arbitration Rules of the Singapore International Arbitration Centre (2016).

124 ibid 435. See, for example, Rule 19 of the Arbitration Rules of the Singapore International Arbitration Centre (2016).

heralded for its important consequences in international arbitration,[125] chief of which is determining the extent of a national court's intervention and the level of deference that it will accord the final award.[126] A tribunal's decision on jurisdiction is subject to *de novo* independent review by the courts, while a tribunal's decision on admissibility is not.[127]

Secondly, the Dichotomy serves to determine practical matters such as who should bear the burden of raising the objection.[128] For instance, a tribunal can review its jurisdiction *proprio motu*,[129] but not the admissibility of a claim which is instead raised by parties.[130] Thirdly, labelling an objection as jurisdictional or admissibility implicates the *res judicata* effect of a tribunal's ruling.[131] A tribunal's decision of lack of jurisdiction carries a *res judicata* effect on the same tribunal, while a ruling of inadmissibility does not invariably bar rehearing of the same claim in the future.[132] In the latter situation where the plaintiff sues too early,[133] the tribunal may find the claim temporarily inadmissible,[134] and stay the proceedings until the relevant admissibility conditions are satisfied.[135]

IV. Conclusion

Ultimately, *BTN* is but one of the many recent decisions where the Singapore courts' stance is made clear in no uncertain terms – the 'tribunal versus claim' test and the dichotomy between jurisdiction and admissibility is here to stay. This pragmatic approach may be lauded for its certainty and efficacy, though the dichotomy is also imperfect in lacking the flexibility that other more open-textured approaches may offer.[136] It appears that the Singapore courts are stuck between a rock and a hard place insofar as alternatives would pose their unique challenges.

125 Walters, 'Fitting a Square Peg into a Round Hole' (n 13) 662; Nota, 'International Arbitration' (n 44) 32.

126 Obamuroh, 'Jurisdiction and Admissibility' (n 9) 375; Walters, 'Fitting a Square Peg into a Round Hole' (n 13) 662.

127 *BTN* (n 2) [66] and [71]. See also N Blackaby, C Partasides QC, Hunter, *Redfern and Hunter on International Arbitration* (n 50) [5.112]; Santacroce, 'Navigating the Troubled Waters Between Jurisdiction and Admissibility' (n 5) 540–541; Tweeddale, 'Jurisdiction and Admissibility in Dispute Resolution Clauses' (n 7) 14.

128 Walters, 'Fitting a Square Peg into a Round Hole' (n 13) 662.

129 Santacroce, 'Navigating the Troubled Waters Between Jurisdiction and Admissibility' (n 5) 551; See, *eg*, Decision 4A_618/2019 of the Swiss Supreme Court, where the court held that the arbitrator was entitled to investigate jurisdiction of its own volition.

130 Yas Banifatemi, 'Chapter 1' (n 8) 19; Nota, 'International Arbitration' (n 44) 32. See also Chittharanjan F. Amerasinghe, *Jurisdiction of International Tribunals* (The Hague, London, New York: Kluwer Law International, 2003) 286; Santacroce, 'Navigating the Troubled Waters Between Jurisdiction and Admissibility' (n 5) 551.

131 Walters, 'Fitting a Square Peg into a Round Hole' (n 13) 662.

132 Obamuroh, 'Jurisdiction and Admissibility' (n 9) 377.

133 *BTN* (n 2) [70]; Paulsson, 'Jurisdiction and Admissibility' (n 6) 616.

134 Santacroce, 'Navigating the Troubled Waters Between Jurisdiction and Admissibility' (n 5) 551.

135 ibid.

136 Apart from Hwang's test, there exists the 'presumed party intentions' test and the 'draftsman' test. See Iris Ng, 'Jurisdiction or Admissibility? The Status of Time Bars Under Singapore Arbitration Law' (Kluwer Arbitration Blog, 7 August 2020) <http://arbitrationblog.kluwerarbitration.com/2020/08/07/jurisdiction-or-admissibility-the-status-of-time-bars-under-singapore-arbitration-law/> accessed 22 November 2021.

In striking the difficult balance between certainty and flexibility, it is hoped that Singapore law will develop in a manner which reduces, or hopefully, eradicates, the twilight zone.