ISSN: 2753-5746

CAMBRIDGE LAW REVIEW

VOLUME VII
ISSUE 1
SPRING 2022

Editors-In-Chief

Andreas Samartzis and Alec Thompson

Editor-in-Chief's Introduction to the Spring Issue of Volume VII of the Cambridge Law Review

It is with great pleasure that I present the Spring Issue of Volume VII of the Cambridge Law Review. This has been another busy period for the journal. We received an impressive number of submissions, many of which demonstrated the combination of originality, research, and attention to detail that make for a thought-provoking read. We also further developed our cooperation with previously established partners and forged a new partnership with De Jure Journal, a student-run publication from the University of Athens.

Volume VII, Issue 1 covers topics from a wide range of areas of law, including comparative constitutional law, law and technology, law and finance, labour law, and legal theory. The issue opens with a comparative analysis of proportionality review in the adjudication of socio-economic rights under English and Hong Kong law. In "Towards a Clearer Expression of the Internal Points of View of Judges in Socio-Economic Rights Adjudication: Lessons from English and Hong Kong Law", authors Thomas Yeon and Benny Chung argue for a more transparent exposition of judges' internal points of view in applying the various stages of proportionality review. This approach could contribute to a judgment's legitimacy by explaining why competing factors relevant to a proportionality assessment are treated in in a certain a manner despite the existence of other viable alternatives.

In "Jumping into the SPAC Race: Protecting the UK Retail Investor", Akansha Maria Paul analyses the issues arising from the recent interest in investment in special purpose acquisition companies (SPACs). Despite their relatively low risk in volatile market conditions, the increasing numbers of retail investors interested in SPACs have given cause for regulatory scrutiny and securities litigation in jurisdictions such as the United States. Contrary to this trend, the UK's Financial Conduct Authority has recently relaxed its Listing Rules to create a more friendly environment for investment in SPACs. The author explores the benefits and shortcomings of the adopted approach.

In the issue's third article, "Tethering the Crypto-Asset Market: The Regulation Of Stablecoins In The European Union And United States", Desiree van Iersel examines the proposals of the European Union and United States for

the regulation of crypto-currencies that are made to maintain a relatively stable value (Stablecoins). The author focuses on the success of these proposals in balancing two factors: the need to encourage innovation, on the one hand, and that of consumer protection, on the other. She concludes that neither proposal fully grasps the complexities of the relevant technology and the continuing evolution of Stablecoins, while offering an adequate level of consumer protection.

Mary Ppasiou offers a comparative analysis of Himalaya clauses in English and Canadian shipping law. Himalaya clauses are provisions in a contract of carriage that extend the carrier's exemptions of limitations of liability under the contract to third parties engaged by the carrier for the performance of the contract. The article "Finding a Home: The Development of Himalaya Clauses in England and Canada" examines how the enforcement of Himalaya clauses has been reconciled with the rule of privity in English and Canadian law. The need to balance principle with commercial practicality has historically resulted in a patchwork of valid legal bases for Himalaya clauses. As the author argues, it is unlikely that the current legal tests for the validity of Himalaya clauses in the two jurisdictions will be the end of the quest for their firm grounding in legal doctrine.

Samson Obiora engages in a critique of Nigeria's collective bargaining framework vis-à-vis the benchmark labour standards of the International Labour Organisation in "Collective Bargaining Trends in Nigeria – Living up to the International Labour Organisation (ILO) Standards?". This critique is informed and enriched by a comparison of the relevant legal framework with the jurisdictions of South Africa and the United Kingdom. The authors calls for reform of the Nigerian legal framework on collective bargaining, with particular emphasis on the right to bargain and the right to organise.

Turning the legal theory, John Choi argues for a re-evaluation of the aims of contract law. Contract law, he maintains, can and should have a more ambitious aim of promoting distributive justice by ensuring that private transactions achieve a fair distribution of wealth. The author's strategy to arrive at this conclusion is to demonstrate its congruence with existing contract rules and to deconstruct possible objections to it. After this intriguing defence of a distributive function for contract law, the author provides an outline of this approach's implications of the interpretation of contract.

The issue closes with a case comment by Samuel Willis on *R (Open Rights Group and the3million) v Secretary of State for the Home Department and Others* [2021] EWCA Civ 1573 (*Open Rights Group (No 2)*. The author argues persuasively that the case enriches legal doctrine on the scope of UK courts' discretion to suspend the effect of public law remedies within the sphere of retained EU law in three respects. Firstly, it identifies an anterior question to be answered when English and Welsh courts are called upon to enforce a rule of retained EU law–namely, whether the rule of law in issue was capable of translation into English and Welsh

law. Secondly, the judgment appears to model, if not explicitly identify, the correct approach to answering that question. Finally, the judgment provides guidance for the exercise of the discretion.

I wish to thank our team of Senior, Associate, and International Editors for their work and dedication during this period. I would also like to express my gratitude to the Honorary Board for their invaluable guidance. I look forward to presenting the Autumn Issue which will be published later in the year.

Andreas Samartzis
Editor-in-Chief

Cambridge Law Review
Editorial Board

Table of Contents

Cambridge Law Review (2022) Vol VII, Issue 1, 1–32

Towards a Clearer Expression of the Internal Points of View of Judges in Socio-Economic Rights Adjudication: Lessons from English and Hong Kong Law

THOMAS YEON* AND BENNY CHUNG**

ABSTRACT

Clear and well-reasoned judgments are key to the development of healthy respect for the judiciary, for they provide practitioners and the public at large with opportunities to understand the fundamental rationales that shape the outcome of cases. This ideal, however, may sometimes come into conflict with robust protection of human rights by way of adopting a stringent standard of review despite the existence of factors suggesting that a more relaxed standard ought to be adopted. In this paper, the approaches to the proportionality test in courts in Hong Kong and the United Kingdom will be unpacked and analysed comparatively. It will be demonstrated and argued that it is essential for judges to spell out their internal legal point of view in the most crystal-clear sense, for the explication of a standard of review in proportionality adjudication necessarily involves two things: (a) explaining why competing factors relevant to the choice of such standard are treated in a certain a manner despite the existence of other viable alternatives; and (b) making a judge's perspectives and reasoning as accessible as possible to the public.

Keywords: socio-economic rights; discrimination law; comparative law; analytical jurisprudence; adjudication

* LLB (Durham); LLM (LSE).

** BA (Berkley); JD (CUHK); BCL (Oxon).

This paper is inspired by our work and discussions with Mr Justice Bokhary NPJ of the Hong Kong Court of Final Appeal and the Hon Geoffrey Ma, whose comments and guidance in our process of drafting this paper we have greatly benefited from and are very grateful for. We would like to express our thanks to the anonymous reviewers for their meticulous comments. The usual disclaimer applies.

I. INTRODUCTION

As Lord Macmillan, a former member of the House of Lords, once cautioned: the object of a reasoned judgment "is not only to do but to seem to do justice".[1] The way in which decisions are written represents choices made by the judge as to how he wishes to justify his ruling.[2] This reminder is particularly apt in the realm of public law litigation, for it often engages controversial issues where the adjudicating court is faced with views diametrically opposed to one another, all of which may appear entirely reasonable. These analytical tensions are best illustrated by the focus of this paper—proportionality adjudication in socio-economic rights and matters concerning entitlements to social welfare and housing. This area is chosen because there exist variable intensities of review (under the third step of the test) and an overall balancing exercise (under the final step of the test) between rights and interests which are often diametrically opposed to one another.[3] These tensions also underpin the broader question this paper seeks to answer: in proportionality adjudication, where there exist multiple choices all of which appear reasonably arguable yet lead to diametrically opposed conclusions, how should a judge choose among the rival options?

This paper argues that since clear and effective communication with the relevant audience is key in rendering a defensible judgment, it is pertinent for judges to make their internal point of view (IPV) of the law—an internal legal point of view (ILPV)—accessible to the general public. The paper proceeds in three sections. First, it explicates the analytical intricacies underpinning a judge's reasoning process in making choices in a proportionality assessment, focusing on the third and fourth steps of the test. Second, it analyses the issues sketched therein through the lens of HLA Hart's IPV—as to how judges can express their preferences and choices in proportionality adjudication. Instead of offering a schema of adjudication or defending a normative understanding of it, the aim of this section is more modest: it seeks to explicate the analytical issues underpinning the practical selection of interpretive and adjudicative choices and the communication of legal reasoning in the process of delivering a judgment. This also avoids over-theorising the problems, which might otherwise result in limited

[1] Lord Macmillan, 'The Writing of Judgments' (1948) 26 Canadian Bar Review 491, 491.

[2] Tayla Steiner, Andrej Lang and Mordechai Kremnitzer, 'Comparative and Empirical Insights into Judicial Practice: Towards an Integrative Model of Proportionality' in Mordechai Kremnitzer, Tayla Steiner and Andrej Lang (eds), *Proportionality in Action: Comparative and Empirical Perspectives on the Judicial Practice* (CUP 2020) 542, 546.

[3] Meghan Campbell, 'The Austerity of Lone Motherhood: Discrimination Law and Benefit Reform' (2021) 41(4) Oxford Journal of Legal Studies 1197.

practical impacts being explicated.[4] Third, the paper will analyse comparatively these issues in Hong Kong and United Kingdom socio-economic rights cases involving discrimination claims. In these cases, the factual and legal factors in a proportionality assessment point towards diametrically different options over the standard of scrutiny (third step) and outcome of the balancing exercise (fourth step). In doing so, this paper will offer some preliminary thoughts towards improving the analytical coherence and clarity on the part of judges when it comes to making choices in a proportionality test.

Before proceeding further, it is helpful to first explain why a comparative exercise between Hong Kong and the United Kingdom is of value. Two reasons may be offered. The first, and more general, explanation is that comparative legal analysis allows practitioners and judges (and of course, academics) to engage with foreign judgments critically and directly. It involves "the recognition that different judiciaries may differ about the resolution of particular classes of legal problems." The analytical dimension of comparative legal analysis helps to avoid practitioners from "*bricolage* — rummaging around just about anywhere for materials which might support particular arguments."[5] The second, and more practical, reason is that the well-known four-stage proportionality test applies in both Hong Kong law and English law,[6] with Hong Kong courts frequently citing English courts (as will be demonstrated below). This makes their adjudicative approaches to its components worthy of comparison[7]. While it is true that the role played by the third step differs slightly between the two jurisdictions,[8] it remains the case that they employ the same notion in expressing the nature and various standards of review commonly seen in the third and fourth steps of the proportionality test. This means that—after all—the audiences in the respective jurisdictions face very similar expressions from the courts as to the outcome of a proportionality exercise and the reason(s) behind it. It is these similar expressions from the courts and reception by the readers that make the two jurisdictions worthy of comparison.

4 James W Harris, *Law and Legal Science* (Clarendon Press 1979) 103; see also more generally Maurice Sunkin, 'The Functionalist Style in Public Law' (2005) 55 University of Toronto Law Journal 361.

5 Robert French AC, "The Globalisation of Public Law: A Quilting of Legalities" in Mark Elliott, Jason NE Varuhas and Shona Wilson Stark (eds), *The Unity of Public Law? Doctrinal, Theoretical and Comparative Perspectives* (Oxford; Hart Publishing, 2018) 231 at 232.

6 *Bank Mellat v Her Majesty's Treasury (No.2)* [2013] UKSC 39, [2014] AC 700; *Hysan Development Ltd v Town Planning Board* (2016) 19 HKCFAR 372;.

7 Thomas Yeon and Trevor Wan, 'Comparative Constitutional and Administrative Law in Hong Kong: In Search for Coherence' [2021] Public Law 261, 268–270.

8 See text to n 16 below.

II. Issues of Clarity in Proportionality Reasoning in Socio-Economic Rights Adjudication

Having set out the comparative basis of the article, it can now turn to the proportionality test as applied in practice. First, the impugned measure must pursue a legitimate aim. Second, the impugned measure must be rationally connected to the legitimate aim. Third, the impugned measure must be no more than necessary for the purpose of achieving the impugned aim. Last, a balance must be struck between the inroads against the relevant human right infringed by the impugned measure and the legitimate aim it sets out to achieve. As an "orderly process of decision-making",[9] it provides a communicable framework to the audiences explaining, amongst others, the legal standard adopted for scrutinising its compatibility with human rights, the reasons behind the standard adopted, and how (if at all) the measure rationally connects to a human right said to be infringed by it.[10] Such communication is particularly important for the third and fourth steps since they "inevitably overlap"[11] with one another in sketching the analytical tensions gravitating a judge towards one conclusion or another as to the proportionality of the measure in question.

The intelligibility of proportionality adjudication in spelling out the quality of a measure said to violate human rights can be found in *Bank Mellat v Her Majesty's Treasury (No.2)* where Lord Reed succinctly explained: "Its attraction *as a heuristic tool* is that, by breaking down an assessment of proportionality into distinct elements, it can clarify different aspects of such an assessment, and make value judgments more explicit."[12] This tool is particularly important for appellate courts (in particular apex courts), since they are tasked to "provide legal certainty, to deliver authoritative statements of the law for the guidance of lower courts, to legitimate specific doctrinal interpretations and extrapolations of the law."[13] That said, a level of generality nevertheless remains appropriate, for the principles and nuances so elucidated govern "many other fact-situations arising in the future

[9] *R (Lord Carlile of Berriew QC) v Secretary of State for the Home Department* [2014] UKSC 60, [2015] AC 945 [89].

[10] Charles-Maxime Panaccio, 'In Defence of Two-Step Balancing and Proportionality in Rights Adjudication' (2011) 24 Canadian Journal of Law and Jurisprudence 109; Kai Möller, 'Balancing and the Structure of Constitutional Rights' (2007) 5 International Journal of Constitutional Law 453.

[11] *Bank Mellat* (n 6) [20].

[12] ibid [74] (emphasis added). Such a need to make value judgments has also been observed in Hong Kong: Johannes Chan, 'Proportionality after *Hysan*: Fair Balance, Manifestly without Reasonable Foundation and *Wednesbury* Unreasonableness' (2019) 49 Hong Kong Law Journal 265, 268.

[13] Peter McCormick, 'The Choral Court: Separate Concurrence and the McLachlin Court, 2000–2004' (2005–2006) 37 Ottawa Law Review 1, 3.

[which] will be governed by that statement of principle".[14] These opposing demands beg the question as to what an approach that can cater to both of them would look like. As will be illustrated below, issues of clarity of reasoning are all the most delicate in the third and fourth steps of a proportionality test.

A. A "SLIDING SCALE" OF REVIEW UNDER THE THIRD STEP

The third step of the proportionality test in both Hong Kong law and the English law admit for a "sliding scale" of review,[15] with the standards "no more than reasonably necessary" (NMRN) and "manifestly without reasonable foundation" (MWRN) on the more stringent and more relaxed ends of the scale respectively. For the MWRN standard, it has been suggested that the usages of the phrase in Hong Kong and the United Kingdom are doctrinally different, with the former jurisdiction using it as an actual limb of the proportionality test and the latter jurisdiction only using it to indicate the appropriate intensity of judicial scrutiny.[16] That said, it will be demonstrated below that such a doctrinal difference in the fields of socio-economic policies and discrimination cases is more apparent than real, and in any event does not dilute the pertinent need for judicial clarity in expressing and explaining the interpretive and adjudicative choices they have made.

It is helpful to turn to the approach in Hong Kong law first, for the operation of the sliding scale under the third step is now settled. This is laid down by Ribeiro PJ of the Hong Kong Court of Final Appeal (HKCFA) in the seminal case of *Hysan Development Co Ltd v Town Planning Board*, which concerned the protection of property rights under Articles 6 and 105 of the Hong Kong Basic Law (BL). After a "magisterial survey of the various [proportionality] doctrines" employed by common law courts,[17] his Lordship established that the choice between the NMRN and MWRF standards revolves around the "margin of discretion" as "determined by factors which affect the proportionality analysis in the circumstances of the particular case."[18] Instead of being disparate standards unrelated to one another, they indicate "positions on a continuous spectrum".[19]

[14] Sir Philip Sales, 'The Common Law: Context and Method' (2019) 135 Law Quarterly Review 47, 55.

[15] *Hysan* (n 6) [108]; *Pham v Secretary of State for the Home Department* [2015] UKSC 19, [2015] 1 WLR 1591 [106].

[16] Kai Yeung Wong, 'An Incomplete Victory: The Implications of *QT v Director of Immigration* for the Protection of Gay Rights in Hong Kong' (2018) 81(5) Modern Law Review 874, 888.

[17] Richard Clayton QC, 'Keeping a sense of proportion: political protest and the Hong Kong courts' [2018] Public Law 375, 378.

[18] *Hysan* (n 6) [106]–[107]. The use of the MWRN standard in socio-economic rights cases trace back to two pre-*Hysan* HKCFA cases, both of which will be examined in more detail in Section IV below.

[19] ibid [122].

The existence of a choice between the NMRN and MWRF standards begs the question behind it: How should such a choice be made? *Hysan* identified two key factors which shape the margin of discretion to be offered to the government: (a) significance and extent of interference with the right in question; and (b) the identity (and special competence, if any) of the decision-maker behind the impugned measure.[20] They were elaborated by the HKCFA in *Kwok Cheuk Kin v Secretary for Constitutional and Mainland Affairs*,[21] where Ma CJ (as the Hon Geoffrey Ma then was) stated that a judge would need to consider three issues in making a choice between the standards: (a) the nature of the right engaged and the degree to which it has been encroached on; (b) the identification of the relevant decision-maker; and (c) the relevance of the margin of discretion that should be given to the decision-maker.[22] Although *Hysan* and *Kwok* concerned contexts of town planning and political speeches, the tests and observations set out therein are of general applicability[23] and, as will be demonstrated below, apply to socio-economic rights cases as well.

For socio-economic rights cases in the United Kingdom, although the four-step proportionality test applies, *R (DA and others) v Secretary of State for Work and Pensions* affirmed that, in terms of the appropriate label to follow in determining the degree of judicial scrutiny under all stages, the government's case would only fail if the applicant can prove in relation to all the four steps that the government's case is MWRF.[24] This means that under the third step, the general applicable standard is MWRF. That said, the MWRF is not a static standard *per se*. In the recent case of *R (SC and others) v Secretary of State for Work and Pensions*,[25] after a meticulous examination of the Strasbourg jurisprudence on the usage of the MWRF standard in social welfare contexts,[26] Lord Reed clarified that:

> [R]ather than trying to arrive at a precise definition of [MWRF], it is more fruitful to focus on the question *whether a wide margin of judgment is appropriate in the light of the circumstances of the case*. The ordinary approach to proportionality gives appropriate weight to the judgment of the primary decision-maker: a degree of weight

[20] ibid [108]–[109].

[21] *Kwok Cheuk Kin v Secretary for Constitutional and Mainland Affairs* (2017) 20 HKCFAR 353.

[22] ibid [38].

[23] Rehan Abeyratne, 'More Structure, More Deference: Proportionality in Hong Kong' in Po Jen Yap (eds), *Proportionality in Asia* (CUP 2019) 25, 40–45.

[24] *R (DA and others) v Secretary of State for Work and Pensions* [2019] UKSC 21, [2019] 1 WLR 3289 [65], [114], [116].

[25] *R (SC and others) v Secretary for Work and Pensions* [2021] UKSC 26, [2021] 3 WLR 428.

[26] These include, among others, *Stec v United Kingdom* (2006) 43 EHRR 47, a Grand Chamber case cited with approval by *Hysan* (n 6) [111] in establishing the "MWRF" limb of the proportionality test in HK.

> which will normally be substantial in fields such as economic and social policy, national security, penal policy, and matters raising sensitive moral or ethical issues.…*the ordinary approach to proportionality will accord the same margin to the decision-maker as the [MWRF] formulation in circumstances where a particularly wide margin is appropriate*.[27]

Lord Reed's more elaborate and nuanced formulation of the approach to the MWRF standard will be further discussed in Section IV below. It suffices to note at this juncture that the MWRF standard, as understood in English law, is not a hard-and-fast standard *per se*, but instead requires multifaceted considerations in determining the appropriate margin of judgement to be accorded to the decision-maker. An illustrative but less obvious example of this can be found in *Re Brewster*,[28] A pre-*SC* judgment. Holding that the MWRF standard applies to an assessment of whether a nomination requirement imposed upon unmarried couples only under a pension scheme violates one's rights under Article 14 of the European Convention on Human Rights (ECHR14) read with Article 1 of Protocol no. 1 to the ECHR (A1P1),[29] Lord Kerr observed that "where the question of impact of a [socio-economic measure] has not been addressed by the government… the imperative for reticence on the part of a court tasked with the duty of reviewing the decision is diminished".[30]

That said, the existence of engagement *per se* is not sufficient, for if the reasons advanced are "proffered in defence of a decision which were not present to the mind of the decision-maker at the time it was made", the standard of scrutiny is likely to be more intense.[31] The converse applies as well, albeit not in an *a fortiori* manner—it remains necessary to determine the weight to be given to the decision-maker's views on a case-by-case basis.[32] These observations are particularly pertinent in the context of socio-economic rights. As Lord Hoffmann observed in *R (Carson) v Secretary of State for Work and Pensions*:[33] "[T]here may be borderline cases where it is not easy to allocate the ground of discrimination"[34] between (a) "discrimination which *prima facie* appear to offend our notions of

[27] ibid [161] (emphases added).
[28] *In re Brewster* [2017] UKSC 8, [2017] 1 WLR 519.
[29] ibid [55].
[30] ibid [64].
[31] ibid [52].
[32] *R (TD) v Secretary of State for Work and Pensions* [2020] EWCA Civ 618 at [54], citing *In re Brewster* (n 30).
[33] *R (Carson) v Secretary for Work and Pensions* [2005] UKHL 37, [2006] 1 AC 173.
[34] ibid [17].

respect due to the individual" and (b) "those which merely require some rational justification".[35]

Therefore, the differences between the MWRF standards in Hong Kong law and English law do not mask the more important and pressing analytical conundrum posed by a "sliding scale" of intensity of review—about the precise degree and method of reasoning which ought to be exhibited thereunder. None of the judgments canvassed above identified any factor to be of overarching or determinative influence for how a standard of scrutiny is to be selected. Nor do they hint to a unifying test for how the relevant factors in a case, if each is seen as inviting a judge to adopt different intensities of scrutiny on the sliding scale, should be weighed against one another. The cases analysed above may be said to identify, broadly, the following three factors as relevant for determining the appropriate standard of scrutiny to be applied: (a) nature of the right engaged; (b) the identity and acts of the decision-maker who enacted the impugned measure, and (c) the identity and acts of the decision-maker who enforced the impugned measure (as alleged to constitute a violation of the right in question).

The lack of a determinative or overarching factor suggests that the third step may be said to be a less-than-absolutely-certain legal test, in the sense that the standard of scrutiny adopted is likely to be unable to cater to all the competing considerations. As the nature of the impugned measure and its operation in practice (which gave rise to the alleged grievance of a claimant) are relevant to the balancing exercise under the fourth step of the proportionality test, the analysis conducted under the third step and conclusions reached therein will have an impact on the balancing exercise conducted under the final stage.

B. Choosing an outcome—as a result of a balancing exercise

The balancing exercise in proportionality analysis is concerned with "the harm caused by limiting the constitutional right", that is, whether the impugned measure excessively burdens the rights of individuals or groups adversely affected by it.[36] Courts in Hong Kong and the United Kingdom must ask whether a fair balance has been struck "between the demands of the general interest of the community and the requirements of the protection of the individual's fundamental rights".[37] The analysis undertaken in this balancing exercise permits a court to "complete" the proportionality analysis by ensuring that no factor of significance has been overlooked in the prior steps, all of which focus more on the

[35] ibid [15].
[36] Aharon Barak, *Proportionality: Constitutional Rights and Their Limitations* (CUP 2012) 344.
[37] *Bank Mellat* (n 6) [70], [73]–[76]; approved in *Hysan* (n 6) [69], [72]–[78].

legitimate aims in question.[38] It provides important room for a court to clarify the community values that it deems to be involved in the case, and in turn make those values transparent.[39] In conducting this assessment, "the legal validity of all of the conflicting principles is kept intact. Their scope is preserved."[40]

In the context of socio-economic rights adjudication in Hong Kong and the United Kingdom, however, this exercise is skewed in both jurisdictions for different reasons. In Hong Kong, Ribeiro PJ suggested that, "in the great majority of cases, [the balancing exercise] would not invalidate a restriction which has satisfied the requirements of the first three stages of proportionality."[41] This is because in such a case, when the impugned measure has passed the first three steps, one would expect that it "internally [reflects] a reasonable balance between the public interest pursued by such laws and the rights of individuals or groups negatively affected by those laws."[42] This observation was followed later in *Kwok*.[43] It, however, appears to sit uneasily with another observation in *Hysan* which was followed in *Kwok* as well:

> Without [the inclusion of the fourth step], the proportionality assessment would be confined to gauging the incursion in relation to its aim. The balancing of societal and individual interests against each other which lies at the heart of any system for the protection of human rights would not be addressed.[44]

These two observations appear to conflict with one another: If the satisfaction of the fourth step is likely to be achieved once the first three steps are satisfied by the government, what is the point of establishing the fourth step in the first place? An answer is available in *Hysan*: "one may exceptionally be faced with a law whose content is such that its application produces extremely unbalanced and unfair results, oppressively imposing excessive burdens on the individuals affected."[45] It remains questionable, however, whether such a possibility can be said to be sufficient for the importance of the right infringed to be adequately reflected, for the assumption of likely satisfaction of the fourth step on the basis of the third step being satisfied by the government risks failing to "examine the

[38] Alec Stone Sweet, 'The Necessity of Balancing: Hong Kong's Flawed Approach to Proportionality, and Why It Matters' (2020) 50 Hong Kong Law Journal 541, 547.
[39] Vicki C Jackson, 'Proportionality and Equality' in Vicki C Jackson and Mark Tushnet (eds), *Proportionality: New Frontiers, New Challenges* (CUP 2017).
[40] Barak (n 36) 346.
[41] *Hysan* (n 6) [73].
[42] ibid.
[43] *Kwok* (n 21) [61(1)].
[44] *Hysan* (n 6) [78]; *Kwok* (n 21) [47].
[45] *Hysan* (n 41).

importance of the right being pleaded."[46] In light of such a skewed focus, the fourth step is more likely than not to be given brief attention only once the government is adjudged to have satisfied the third step of the proportionality test.[47]

The problem of a skewed fourth step may also be found in English law, with the adoption of a MWRF test in the fourth step being broadly reflective of a similar issue. In *DA*, Lord Reed concluded that the MWRF test continues to apply in the fourth step of a proportionality test on a human rights challenge involving socio-economic matters, as to whether a balance has been struck between the A1P1 rights of the aggrieved applicant (read with ECHR14) and the objectives pursued by the impugned measure.[48] Putting aside the issue of the correctness of this approach as a matter of following Strasbourg jurisprudence,[49] the use of the MWRF standard appears to conflict directly with the notion of "balancing" between an individual right and the aims pursued by the impugned measure. Regrettably, this implication of this approach was not clarified by Lord Reed in *SC*, which focused more on the nature of the MWRF standard as applied under the third step.[50]

These skewed foci raise issues of reasoning and presentation about how a defensible judicial articulation of the ILPVs in making a choice between the competing rights and interests should be made, and how the articulation of ILPVs in the third step affect those to be made in the balancing exercise. A robust understanding in that regard is desirable since, as Sir Philip Sales (as Lord Sales then was) astutely argued, common law adjudication is supported by "its sensitivity to the particular facts of individual cases and from *being able to make localised accommodating of competing values*."[51]

III. The Internal (Legal) Point of View of Judges

Having sketched the analytical conundrums and potential disjunctions inherent in the third and fourth steps of a proportionality test, this paper turns to examine them from the lens of analytical jurisprudence—how these issues ought to be practically resolved in light of the established components of the proportionality test. It will be argued below that the ILPV of judges can be of credible assistance

[46] Sweet (n 38) 554.
[47] Abeyratne (n 23) 53–57; Chan (n 12) 270–271.
[48] See text to n 24 above.
[49] On this point, see Jed Meers, 'Problems with the "manifestly without reasonable foundation" test' (2020) 27(1) Journal of Social Security Law 12, 15–17.
[50] See text to n 27 above.
[51] Sales (n 14) 58 (emphasis added).

in clarifying and making sense of these conundrums, and in turn enable judges to develop clearer processes of reasoning.

A. Introducing the internal (legal) point of view

The concept of an IPV was first articulated by HLA Hart, who defined it as follows: "the view of those who do not merely record and predict behaviour conforming to rules, but *use the rules as standards for the appraisal of* […] *others' behaviour*."[52]

Championed as a "decisive advance for analytical jurisprudence",[53] the IPV is a "hermeneutic concept"[54] that is helpful when one is not only observing the thoughts of actors in a legal system, but also articulating how an individual reasons and operates.[55] This includes judges as well. For example, for a judge to say "it is the law that…" would signify his acceptance of the statement or proposition referred to which is labelled as "the law". This in turn "manifest[s] [his] own acceptance of them as guiding rules", illustrating his IPV vis-à-vis the nature of the legal statement in question.[56] As Scott Shapiro observes, in articulating the IPV, Hart intends to "render the thoughts and discourse of legal actors comprehensive. The [IPV]… [explains a legal activity's] very intelligibility."[57] Considering this methodological injunction in the context of a judge's reasoning process, the articulation of his IPV on the rules and of principles of law he faces—the ILPV—may therefore be of hermeneutic assistance in explicating the reasons and purposes for which certain interpretive and-or adjudicative choices were made.[58] It seeks to make sense and rationalise the adjudicative choices made by a judge when he faces legal rules and principles for which, there are respective reasonably arguable cases that they should be applied, despite gravitating a court towards diametrically opposed outcomes. More importantly, the use of the ILPV as a methodological injunction to illustrate a judge's reasoning process and the choices[59] made therein shifts "attention away from philosophical abstractions

52 HLA Hart, *The Concept of Law* (3rd edn, Clarendon Press 2012) 98 (italics and underline emphasis original; italics emphasis added).

53 Neil MacCormick, *H.L.A. Hart* (Stanford University Press 1981) 32.

54 Brian Leiter, 'Beyond the Hart/Dworkin Debate: The Methodology Problem in Jurisprudence' (2003) 48 American Journal of Jurisprudence 17, 41.

55 Hart (n 52) 88–91.

56 ibid 102, 117.

57 Scott Shapiro, 'What is the Internal Point of View?' (2006) 75(3) Fordham Law Review 1157, 1166; Jules L Coleman and Brian Leiter, 'Legal Positivism' in Dennis Patterson (ed), *A Companion to Philosophy of Law and Legal Philosophy* (Blackwell Publishing 1996) 241, 247.

58 William Lucy, *Understanding and Explaining Adjudication* (Clarendon Press 1999) 58–62.

59 Shapiro (n 57) 1167.

toward a more practical view of law-as-social-activity".[60] The importance of adopting such a practical view, bearing in mind the social dimensions of legal reasoning and application of legal principles, is vividly illustrated by Lord Sales JSC extra-judicially:

> The common law gains from its sensitivity to the particular facts of individual cases and from being able to make localised accommodations of competing values. It can reflect forms of social knowledge embedded in practical experience and local understandings of how to do things well, which may be hard to articulate and state in abstract terms. This sort of knowledge may be ignored where the state tries to proceed by laying down abstract general rules in advance, potentially at great cost to society.[61]

The IPV has been embraced by prominent Anglo-American jurists whose accounts of jurisprudence have an emphasis on adjudication, albeit not in language identical to the preliminary sketch of ILPV in the foregoing paragraph.[62] For Neil MacCormick, the IPV helpfully demonstrates that "the acceptance of rules is not unreasoned, though indeed different people may reason differently for acceptance of the same rule".[63] In considering competing propositions of law, it is appropriate to resort to arguments focusing on the consequence of a certain proposition, such arguments being "essentially evaluative and therefore in some degree subjective".[64] Such a reasoning process does not entail absolute judicial discretion,[65] but instead judges only have "quite restricted freedom of manoeuvre as they try to work through to a reasonably *justifiable* conclusion *justified as a conclusion of law* in the case seen as a *legal case*".[66] Similarly, Ronald Dworkin notes,

60 Allan Hutchinson, 'A Postmodern's Hart: Taking Rules Sceptically' (1995) 58 Modern Law Review 788, 798. In *The Concept of Law*, Hart suggests that the explication of the IPV of a person is not only confined to judges, but also to "any educated man": Hart (n 52) 6.

61 Sales (n 51). As to how the inherently public and justificatory aspect of the common law judicial decision-making both predetermines the form of a judgment must take and creates the community or communities which will evaluate and validate the judgment's legal status, see Douglas E Edlin, *Common Law Judging: Subjectivity, Impartiality and the Making of Law* (University of Michigan Press 2016) 50.

62 Lucy (n 58) 70–75.

63 Neil MacCormick, *Legal Reasoning and Legal Theory* (OUP 1978) 64.

64 ibid 105.

65 In this regard, it is helpful to note that Hart's invocation of "strong discretion" to justify the invocation of judicial power in light of competing interpretations, sketched by the IPV of a judge, is generally seen as ill-substantiated and should at best be "viewed as a dogma": AWB Simpson, 'Common Law and Legal Theory' in AWB Simpson (eds), *Oxford Essays in Jurisprudence: Second Series* (Clarendon Press 1973) 80.

66 Neil MacCormick, *Rhetoric and the Rule of Law: A Theory of Legal Reasoning* (OUP 2005) 147 (emphasis added).

from the perspective of interpretation, that in seeking to give an account of social practice, one may choose to report only "the various opinions different individuals in the community have about what the practice demands".[67] In choosing between interpretations on a legal principle, judges should attempt to find the "best" answer portraying the law "in its best light".[68]

The foregoing juristic illustrations for the limits and communication of judicial reasoning suggest that instead of being confined to theoretical discussions in analytical jurisprudence, the ILPV can be instrumental in illustrating how a judge reasons through multiple interpretations of a legal principle. They also demonstrate how, despite the variety of options available, a particular conclusion is reached. That said, it remains that such a choice made is an individual and subjective choice. It is subjective because it is up to the judge himself to decide which option to follow, instead of being bound by, for example, *stare decisis* or the decisions of apex courts. The exercise of choice in adjudication goes to "the manner in which [one] should understand the judicial failure to admit whilst adjudicating that [judges'] decisions are, or at least on some occasions, the result of individual choices rather than the application of pre-existing standards".[69] As a "permits-based (and not chance-based)" function, this choice "cannot be arbitrary".[70] This demand underscores the epicentre of the analytical conundrums sketched since the beginning of Section II—about the justification of opting for one choice over another in public law adjudication, and the quality of justification advanced. More specifically, the question can be posed as follows: How can the ILPV of judges provide a satisfactory answer to the conundrums and communication deficiencies generated by the need to make interpretive and adjudicative choices in public law adjudication?

It is important to first bear in mind that the ILPV is an IPV about law, meaning that it is about articulating the reasons for opting for a relevant and persuasive authority over another. This means that, instead of jumping directly to asserting that one option is to be adopted, the first step in explicating this ILPV should be to elaborate on the relevance and weight to be accorded to each factor relevant to the aforesaid judicial need to make a choice. This exercise would involve considering how the facts fit within the relevant framework of adjudication. It provides a foundation for the hermeneutic basis of the ILPV by clearly demonstrating the basis on which the relevant options are legally rooted. In setting out and elaborating upon these factors, judges do not act in an unconstrained

67 Ronald Dworkin, *Law's Empire* (Harvard University Press 1986) 64.
68 ibid 255.
69 Richard Nobles and David Schiff, 'Why Do Judges Talk the Way They Do?' (2009) International Journal of Law in Context 25, 40.
70 Joe McIntyre, *The Judicial Function: Fundamental Principles of Contemporary Judging* (Springer 2019) 93.

manner. Instead, the articulation of these factors can and is likely to be shaped by, amongst others, "the legal language, the corpus of legal rules, concepts, principles, and ideas, legal processes and practices, hierarchical legal institutions, [and] the craft of lawyering".[71] They can "[stabilise] legal meaning and [provide] restraint on the influence of subjective views" of a judge.[72]

After setting out the nature of the factors relevant to the choice a judge needs to make, he would then proceed to explain how they contribute to the decision which he has reached. This involves, of course, explaining why a particular conclusion is reached. If, despite the existence of multiple authorities (all of which can be reasonably argued as relevant and applicable to the choice facing a judge), he simply asserts that one is selected, this clearly falls short of the ILPV's analytical and communication demands. Proffering positive reasons justifying the authority selected would be a good start, but there is no logical guarantee that they can also serve as answers for rejecting an authority which is not selected (and followed). In such a scenario, it cannot be said that the option that is not selected is wrong or inapplicable, for it is not the case that the aforesaid option is wrong as a matter of law. It is simply disapproved because of a judge's subjective choice not to apply it in reaching a conclusion on the relevant issue. Therefore, in articulating the ILPV, instead of trying to sketch the line of reasoning or interpretation so selected as analytically watertight as a matter of law, it would be more practical and indeed more appropriate for the articulation of the ILPV in that regard to be characterised as a matter of justifiability or defensibility. The illustration should focus on showing that the choice made is justifiable or defensible, not that it must be correct.[73]

The qualification of the level of reasoning required as one of justifiability but not offering an absolute answer is critical for acknowledging that the choice is a subjective one—a critical component of the articulation of the ILPV. Both the IPV and its hermeneutic application in the context of adjudication demonstrate that it would be unsatisfactory for a judge to rid himself of subjectivity at the expense of clarity of analysis and communication of his decision to an audience. Instead, subjectivity can be embraced, especially when it comes to instances where there is no clear answer to the relevance, weight and application of legal principles and facts.[74] Illustrating the consideration of competing factors and conclusions, in

71 Brian Tamantha, *Laws as a Means to an End: Threat to the Rule of Law* (CUP 2006) 239.

72 ibid 240.

73 John Bell, 'Discretionary Decision-Making: A Jurisprudential View' in Keith Hawkins (ed), *The Uses of Discretion* (Clarendon Press 1992) 89, 106. Indeed, reasoned judgment is essential to the common law method of judging, allowing a judgment (and points of decision therein) and the process(es) of reaching so be "debated, attacked and defended": David L Shapiro, 'The Defense of Judicial Candour' (1987) 100 Harvard Law Review 731, 737.

74 Peter Cane, 'Consequences in Judicial Reasoning' in Jeremy Horder (ed), *Oxford Essays in Jurisprudence (Fourth Series)* (OUP 2000) 41.

particular those that are unfavourably treated, is conducive to demonstrating to the public that the court has (a) adopted a defensible criterion or merit, (b) demonstrated adequately the factual and legal matrices supporting the application of that norm; and (c) reached the result the judge deems most appropriate.[75]

B. THE ILPV OF JUDGES AND THE ASSOCIATED ADJUDICATIVE CONUNDRUMS IN THE CONTEXT OF THE THIRD AND FOURTH STEPS OF PROPORTIONALITY ANALYSIS: PROVISIONAL ILLUSTRATIONS

The foregoing sections of this paper advance two overarching points. First, adjudicative choices permeate the third and fourth steps of proportionality adjudication, for they involve selecting multiple viable propositions of law and interpretations leading to different (and sometimes diametrically opposed) outcomes. Second, the ILPV can serve as a powerful analytical and communication tool in demonstrating the various interpretive exercises and choices made in the course of adjudication and delivering a judgment, particularly when the answer is not clear-cut from the available authorities and facts.

Placing the ILPV in a proportionality framework demonstrates the analytical and communication difficulties that judges face when conducting proportionality analyses. First, instead of being confined to questions of law only, the formation and the operation of the internal point of view cannot be divorced from questions of facts. This is because under the third and fourth steps of a proportionality test, courts may be required to make specific fact-findings.[76] For example, a court may be tasked to make a finding on the precise degree of impact that the infringement of the right (as established) has on the aggrieved applicant, or the actual reach of the identified objectives of the impugned measure in practice. The involvement of fact-finding exercises adds a further complexity to our problem: the presentation of the ILPV would also have to demonstrate how the relevant fact-finding exercises would impact the articulation of the ILPV in the first place. It is more often than not that instead of having a stand-out answer, adjudicative choices hereunder involve "choosing between *several* legitimate options".[77]

Second, and as illustrated above, in exercising adjudicative discretion in a proportionality test, it is likely that the judge's decision may be infused with subjective views. This may encompass, amongst others, his views on the role of a

[75] Melvin Eisenberg, 'The Principles of Legal Reasoning in the Common Law' in Douglas E Edlin (ed), *Common Law Theory* (CUP 2007) 81, 94.

[76] *R (Kiarie) v Secretary of State for the Home Department* [2017] UKSC 42, [2017] 1 WLR 2380 [46]–[47].

[77] Aharon Barak, 'On Judging' in Martin Scheinin, Helle Krune and Marina Aksenova (eds), *Judges as Guardians of Constitutionalism and Human Rights* (Edward Elgar 2016) 27, 29–30 (emphasis added).

court vis-à-vis the other branches of government and thoughts on the discretion as to the appropriate development of the area of law in question.[78] Appeals to arguments of, for example, common sense should not be seen as of assistance. While such an argument would appear to have strong logical force that is accessible and-or easily acceptable by the audiences of a judgment, such appeals "do not define a distinct method of legal reasoning that can make a plausible claim to intellectual rigour".[79] Similarly, the defensibility of choices as manifested in a judge's ILPV is uncertain in the final step, where "rights do not… enjoy any special or elevated status over public interests, but rather *operate on the same plane as policy considerations*."[80] Since the right infringed by the impugned measure and the legitimate objectives that it pursues are compared head-on with one another, clear articulation of the weight (and importance) to be accorded to each factor contributing to the two aforesaid items of consideration becomes all the more indispensable. The explication of value judgments inherent in the balancing exercise, in light of the aforementioned articulative task, means that the heuristic nature of proportionality analysis is likely to be shaped considerably by such an exercise.

Armed with the foregoing observations, this paper will now turn to the sources of the analytical and communicative conundrums sketched above: (a) the "sliding scale" of intensity of review under the third step; and (b) the balancing exercise and holistic evaluation of the right infringed and the legitimate objectives pursued. Although the cases under examination concern socio-economic policies where discrimination claims are alleged, it will be illustrated that the analytical and communicative challenges that the ILPV can tackle extends to proportionality adjudication in general as well, especially for the final balancing exercise.

[78] Aharon Barak, *Judicial Discretion* (Yale University Press 1989) 72 ("Judicial decision reaches its peak when the judge strikes a balance between competing principles, according to their weight and their strength at the point of confrontation").

[79] Richard Posner, *How Judges Think* (Harvard University Press 2008) 117.

[80] Kai Möller, *The Global Model of Constitutional Rights* (OUP 2012) 10–15 (emphasis added).

IV. No More Than (Reasonably) Necessary: A Choice of Standards

A. Different factors gravitating towards different directions and their relationship with one another

(i) Hong Kong Law: A Straightforward Attachment of Weight to "Suspect" Grounds

Dissecting and evaluating the reasoning process of courts in socio-economic rights cases involving discrimination claims require a brief detour to two pre-*Hysan* HKCFA cases. First, in *Fok Chun Wa v Hospital Authority*,[81] the applicant challenged the respondent's policy of imposing higher fees for non-Hong Kong residents (compared to Hong Kong residents) giving birth in public hospitals on the basis of an infringement of her right to equality (protected under Article 22 of the BL (BL25) Article 22 of the Hong Kong Bill of Rights (BOR22)).[82] Delivering a unanimous judgment, Ma CJ observed that the involvement of socio-economic policies *per se* "does not lead to the consequence that [courts] will not be vigilant when it is appropriate to do so or that the authorities have some sort of carte blanche".[83] The need to attach weight to the identity of the decision-maker and its competence in socio-economic matters does not require uncritical deference towards any decision rendered. Where the unequal treatment "strikes at the heart of core-values relating to personal or human characteristics… the courts would extremely rarely (if at all) find this acceptable [because] these characteristics involve the respect and dignity that society accords a human being".[84] Barring an immediate jump to a conclusion that a more stringent standard of scrutiny should always be applied,[85] it is clear that more weight would be accorded to the nature of the infringed right.

The second case is *Kong Yunming v Director of Social Welfare*,[86] which concerned a challenge brought by a new immigrant from mainland China against the constitutionality of the seven-year residence requirement for receiving welfare assistance under the Comprehensive Social Security Assistance scheme. Following

[81] *Fok Chun Wa v Hospital Authority* (2012) 15 HKCFAR 409.

[82] Hong Kong has a "tripartite" regime of rights-protection. The BL enshrines various fundamental rights and freedoms. Article 39 of the BL provides that provisions of the International Covenant on Civil and Political Rights remain in force and implemented through the BOR, as contained in the Hong Kong Bill of Rights Ordinance (Cap.383).

[83] *Fok* (n 81) [77].

[84] ibid.

[85] CL Lim, 'Judicial Rhetoric of a Liberal Polity: HK, 1997–2012' in Rehan Abeyratne and Iddo Porat (eds), *Towering Judges: A Comparative Study of Constitutional Judges* (CUP 2021) 117, 131–132.

[86] *Kong Yunming v Director of Social Welfare* (2013) 16 HKCFAR 950.

Fok, Ribeiro PJ held that "the adoption of a residence requirement as a criterion of eligibility for social welfare benefits… is generally not regarded as engaging any of the inherently suspect grounds".[87] This in turn justifies the general adoption of the MWRF standard in the third step of a proportionality assessment of socio-economic policies (an area where the government enjoys a "wide margin of discretion")[88] but does not engage "inherently suspect grounds" of discrimination (for example, sex or sexual orientation).[89] Once a "suspect ground" of discrimination is engaged, however, it would automatically be accorded more weight.[90] From the lens of ILPV, this entails that the nature of the right is accorded determinative weight in the reasoning process leading towards the application of a more stringent standard of review.

The ostensibly dominant weight accorded to the nature of the infringed right once a "core value" or "suspect" ground is involved continued in two post-*Hysan* HKCFA judgments concerning discriminatory treatment in violation of the applicants' (amongst others) BL25 and BOR22 on the basis of one's sexual orientation. In *QT v Director of Immigration*,[91] a case concerning the Director's refusal to grant the aggrieved applicant a dependent visa on the basis of his same-sex marriage, the court noted that discrimination on a suspect ground is "especially pernicious"[92] because such a ground concerns a "personal characteristic" on the basis of which any differential treatment would be "particularly demeaning for the victim".[93] The same approach is followed in *Leung Chun Kwong v Secretary for Civil Service*,[94] a case concerning the Secretary's decision not to allow the aggrieved applicant, on the basis of their same-sex orientation, to elect for joint assessment of salaries tax with his partner under the Inland Revenue Ordinance (Cap. 112). Both cases placed significant emphasis (following *Fok*, *Kong*, *Hysan*, and *Kwok*) on the need to attach great weight to a suspect ground of discrimination when elaborating on the relevance and nature of the factors as contextualised within the legal and factual matrices of the case. Now recall the three factors which shape the legal basis for adopting a standard of review. It can be observed that the engagement of a "suspect ground" meant that the nature of the right, as a relevant but not determinative factor, was automatically given predominant weight. No comparative exercise has been carried out to illustrate how, let alone explain why, weight is to be given to the identity and competence

87 ibid [42].
88 ibid.
89 ibid [40]. See also nn 84–85 above.
90 ibid. See also Karen Kong, '*Kong Yunming v Director of Social Welfare*: Implications for Law and Policy on Social Welfare' (2014) 44(1) Hong Kong Law Journal 67, 73–74.
91 *QT v Director of Immigration* [2018] HKCFA 28, (2018) 21 HKCFAR 324.
92 ibid [107].
93 ibid citing *Carson* (n 33) [55].
94 *Leung Chun Kwong v Secretary for Civil Service* [2019] HKCFA 19, (2019) 22 HKCFAR 127.

of the decision-maker who enacted and enforced the impugned measure. The leap in reasoning in the string of HKCFA cases demonstrated that, whilst providing a logical basis for discriminatory measures to be subjected to more intense judicial scrutiny, the current formulation works at the expense of clear communication to the public of the precise nature of the interactions between relevant factors.

Despite the apparent emphasis on the nature of the right engaged in socio-economic right cases involving a "suspect" ground of discrimination, the flexible framework set out in *Hysan* and *Kwok* remains able to allow courts to adopt a more nuanced and elaborate approach in instances where the relevant factors gravitate the court towards diametrically opposed outcomes. Two recent decisions of the Hong Kong Court of First Instance (HKCFI) demonstrate this potential. In *Infinger v Hong Kong Public Housing Authority*,[95] a case concerning the infringement of the same-sex applicant's BL 25 and BOR22 rights as a result of different application requirements imposed on opposite-sex and same-sex applicants for public rental housing, Chow J (as Chow JA then was) observed that while the limited supply of such housing may entitle the decision-maker to have "a wide margin of discretion in the performance of its function and responsibility", this factor "should not… be overly emphasised".[96] The "scarcity of public resources" is only one factor to be taken into account in articulating the applicable standard of scrutiny.[97] Similarly, in *Ng Hon Lam Edgar v Hong Kong Housing Authority*,[98] where the aggrieved applicant challenged the respondent's exclusion of the same-sex spouse from owning a Home Ownership Scheme flat from premium-free transfer of title, Chow JA clarified that the "strength of the legitimate aim" does not have a determinative bearing on applicable standard of scrutiny, and nor does the involvement of indirect discrimination necessarily point to a "lower standard or intensity of review".[99] The impugned measure's irrelevance to the allocation of flats *per se* meant that the public resource factor should be given less weight in articulating the appropriate standard of scrutiny.[100]

The approaches in the two HKCFI decisions canvassed in the foregoing paragraph are, compared to the HKCFA's approaches in *Fok*, *Kong*, *QT*, and *Leung*, more elaborate in demonstrating how and why certain degrees of weight are attached to the relevant factors. While it is true that *Infinger* and *Ng* concerned (unlike the HKCFA cases above)[101] scarce public resources on which the executive and/or legislature are usually seen as possessing even greater expertise in making

[95] *Infinger v Hong Kong Public Housing Authority* [2020] HKCFI 329, [2020] 1 HKLRD 1188.
[96] ibid [42].
[97] ibid.
[98] *Ng Hon Lam Edgar v Hong Kong Housing Authority* [2021] HKCFI 1812, [2021] 3 HKLRD 427.
[99] ibid [64(2)]–[64(5)].
[100] ibid [65].
[101] *Infinger* (n 95) [44].

decisions on them, it is clear that the more nuanced articulation of the relevant factors gravitating a court towards adopting standards of review of different stringency provide clearer analytical and communicative basis for justifying the court's eventual choice. It allows a judge to explicate more intelligibly the contribution of the factors towards his reasoning process in the operation of the "sliding scale" of review. The current approaches exhibited by the HKCFA, compared to the HKCFI approaches, provide less helpful hermeneutic bases for communicating to the audience how the opposing (diametrically or not) factors are handled and given consideration.

(ii) Juxtaposing against English Law: Similar Problems, Less Obvious

Sceptics may criticise that, since the majority of the Supreme Court in *DA* has already firmly concluded that MWRF is the applicable standard in decisions concerning socio-economic matters and welfare benefits, any further exploration of the attachment of weights to the MWRF standard would only be a storm in a teacup. Such an argument fails to appreciate the nuances and flexibility involved when a judge applies the standard in practice,[102] for it rests only on the "primarily precedential"[103] rationale of adopting it. It also oversees the analytical conundrums, from the lens of the ILPV, posed by the standard and possible instances of more stringent scrutiny.

The first case where the MWRF standard and the potential for a more stringent standard of scrutiny are discussed in detail is *Humphreys v Revenue and Customs Commissioners*,[104] a case concerning differential treatment between men and women towards child tax credit. In that case, after holding that the standard applies to state benefit cases,[105] Lady Hale cautioned that it does not mean that the impugned measure "should escape careful scrutiny".[106] Similarly, in the latter case of *R (SG) v Secretary of State for Work and Pensions*,[107] a case concerning a cap on welfare benefits for non-working households which is said to amount to unjustifiable gender discrimination under ECHR14 read with A1P1, Lord Reed cautioned that the economic and political judgment involved in the impugned measure contribute to "major implications for public expenditure"; the measure has also been the "subject of full and intense democratic debate".[108] That said, the

[102] Meers (n 49) 17–20.

[103] Mel Cousins, 'The 'benefit cap' and human rights before the UK Supreme Court' (2020) 27(1) Journal of Social Security Law 23, 27.

[104] *Humphreys v Revenue and Customs Commissioners* [2012] UKSC 18, [2012] 1 WLR 1545.

[105] ibid [17]–[19].

[106] ibid [22].

[107] *R (SG) v Secretary of State for Work and Pensions* [2015] UKSC 16, [2016] 1 WLR 1449.

[108] ibid [92]–[95], A similar observation was also made by Lady Hale (dissenting as to the application of the test to the facts): [159].

requirement for giving weighty reasons to justify discrimination is consistent with the *Humphreys* approach.[109] These two cases do not go as far as the analytically questionable immediate jump to a more stringent standard of scrutiny seen in *QT and Leung* in Hong Kong. But, on the other hand, they also do not illustrate fruitfully for how the reasoning process through competing factors is to be conducted. In particular, the focus on the "economic terms" of the measure in *SG* has been suggested to contribute to the disproportionate focus on the institutional identity and competence of the decision-maker.[110]

Returning to *DA*, in which both *Humphreys* and *Carmichael* were cited with approval,[111] Lord Wilson stated that under the MWRF approach, the question is to "[inquire] into the justification of the adverse effects of rules for entitlement to welfare benefits by reference to whether they are [MWRF]".[112] Elaborating on the reasoning process under the MWRF standard, Lord Wilson noted that:

> The rationalisation has to be that, when the state puts forward its reasons for having countenanced the adverse treatment, it establishes justification for it unless the complainant demonstrates that it was [MWRF]. But reference in this context to any burden … is more theoretical than real. The court will proactively examine whether the foundation is reasonable.[113]

Two brief points may be made about this conclusion. First, the "unless" formulation, whilst compatible with precedent, falls short of demonstrating a clear degree of analysis that ought to be demonstrated in relation to the weight to be accorded to competing factors under the test. The question of competing factors remains live, for the MWRF standard has never been a hard-and-fast rule and admits some degree of flexibility. Second, and on a related note, the lack of elaboration as to the assignment of weight to competing factors risks succumbing to the opaqueness in reasoning displayed in *QT* and *Leung*. Both approaches appear to adhere to a jump to a conclusion about the applicable standard of review, despite their observations on the need for more nuanced considerations to the contrary. This neglected need for more nuanced consideration is also supported by Lord Carnwath's observations in the same case, that the submissions were complicated with "conflicting factual and statistical evidence, much of it produced for the first time in this court".[114]

[109] ibid [11]. See also ibid [125] (Lord Carnwath).
[110] Campbell (n 3) 1204.
[111] *DA* (n 24) [61]–[62].
[112] ibid [59].
[113] ibid [66].
[114] ibid [123].

Indeed, post-*DA* decisions continued to illustrate the relevance and pertinence of articulating the weight accorded to each factor deemed relevant to the formulation of the standard of review under the third step. In *R (Joint Council for the Welfare of Immigrants) v Secretary of State for the Home Department*,[115] Hickinbottom LJ cautioned that the reasoning process leading up to the application of the MWRF standard is not "a simple binary question".[116] Instead, the court will examine the basis of the impugned differential treatment (for example, race, nationality, gender), the objective of the impugned measure, and the factual context in which it was enforced.[117] This approach, whilst not elevating the application of the test into "a debate about the precise content of stringency of the MWRF test in a case when it unquestionably applies",[118] illustrates that the MWRF test ought not be treated as a straightforward application exercise. Beneath a general approach that the MWRF test applies in socio-economic matters lies a hybrid of considerations including, but not limited to, the assessment undertaken by the decision-maker(s) in relation to the enactment and enforcement of the impugned measure and the need to give great weight to the engagement of a suspect ground of discrimination.[119] In contrast to the Hong Kong approach sketched above, the engagement of a suspect ground of discrimination does not automatically entail significant weight being given to that factor at the expense of paying insufficient attention to (or failing to articulate so) factors relating to, for example, the institutional competence and expertise of the decision-maker. The English approach therefore enables more room for a wider picture to be painted for a judge's ILPV on how the factors identified to be relevant to the choice of standard of scrutiny are accorded respective weight (if any).

B. Justifying the choice of standard: the reasoning process

Having identified the content of the relevant factors and their significance in the case at hand, a judge would then proceed to justify his choice of the standard of review adopted. As will be demonstrated below, the issue of analytical clarity and expression of choice is of no less importance here, for it rationalises and makes coherent the factors identified in the prior step explained above.

[115] *R (Joint Council for the Welfare of Immigrants) v Secretary of State for the Home Department* [2020] EWCA Civ 542, [2021] 1 WLR 1151.

[116] ibid [136].

[117] ibid.

[118] *R (Delve and another) v Secretary of State for Work and Pensions* [2020] EWCA Civ 1199, [2021] 3 All ER 115 at [90].

[119] *Joint Council* (n 115) [148(iv)].

(i) Hong Kong Law: Moving Away from a Logically Desirable but Analytically Questionable Jump to NMRN

The analytical incoherence generated by the automatic conferral of significant (if not predominant) weight to a "suspect" ground of discrimination in socio-economic rights cases in *QT* and *Leung* extends to the standard-justifying process as well. In both cases, the HKCFA stated unequivocally (albeit in *obiter dicta*) that once an individual is subject to differential treatment on a suspect ground, a court should apply the NMRN standard.[120] These approaches, whilst not necessarily incompatible with the reference to a "sliding scale" of review, exhibit the same weaknesses in reasoning and communication (as explained above) of an insufficient consideration of countervailing factors gravitating a court towards a more relaxed standard of scrutiny. The starting point of affording great weight to a suspect ground has therefore resulted in the application of a more stringent standard of review which, despite being able to offer stronger protection to aggrieved individuals, risks sacrificing clarity in communicating to the audience the steps of analysis adopted by a court when considering the countervailing factors in play.

The interwovenness of the assignment of weight to factors relevant to the operation of the sliding scale and the reasoning process justifying the standard adopted can also be observed in the HKCFI cases examined above. In *Infinger*, after noting the diametrically opposed tensions posed by the engagement of a suspect ground of discrimination on the one hand and the involvement of scarce public resources on the other,[121] Chow J set the applicable standard of review at "somewhere in the middle of the continuous spectrum of reasonableness, and the intensity of review should be set accordingly".[122] In *Ng Hon Lam Edgar*, a case with a fact pattern similar to that of *Infinger*, Chow JA held that the applicable standard of review "should be somewhere between the middle and high end of the intensity of review in the continuous spectrum of reasonableness".[123]

The analytical and communicative ambiguity reflected in the conclusions reached in *Infinger* and *Ng Hon Lam Edgar* illustrates two points. First, the existence of competing factors gravitating a court towards diametrically opposed intensities of scrutiny means that imprecise articulation of the weight to be conferred upon each factor would undermine the legal support those factors may provide for a judge's reasoning process. Once a judge gets off on a murky starting point, the reasoning process of the judge is more likely to be seen as an ill-substantiated and abrupt demonstration of a holistic consideration of the factors by selecting a

[120] *QT* (n 91) [105]–[108]; *Leung* (n 94) [22].
[121] *Infinger* (n 95) [42]–[43].
[122] ibid [44].
[123] ibid [66].

middle point. Both mistakes fall short of the analytical and communicative demands posed by the ILPV. The second point, related to the first, is that in instances where the competing factors are diametrically opposed to one another by themselves (instead of just gravitating towards diametrically opposed outcomes), it is all the more pertinent for a judge to acknowledge the existence of a judgment on his part on the respective weights to be attached to them. By explicating the inescapable need to make, for example, a value judgment in assigning weight to each factor, the heuristic value of the standard of scrutiny can be preserved. This falls short of completing a gapless picture for the application of law to the facts, but it at least preserves the communicative clarity on the part of the judge which, as stressed above, is an indispensable matter that a presentation of reasoning process ought to safeguard jealously.

(ii) English Law: MWRF but Developing a More Nuanced Potential

As argued above, *DA*'s confirmation of the applicability of MWRF in socio-economic policies cases ought not be interpreted as diminishing the need for judges to explicate the factors that may be said to gravitate a court towards different standards of review. The heavy characterisation of economic and political judgments involved in the cases canvassed above in turn affects explanations about why, despite a party's arguments to the contrary, the MWRF standard is adopted. In *R (Delve and another) v Secretary of State for Work and Pensions*, the appellant challenged unsuccessfully the judge's application of the MWRF standard for assessing the equalisation of state pension ages for women with that of men made via revisions to a series of Pensions Acts between 1995 and 2014. On the appropriate scope of deference, the court stated that the Pensions Acts deal with "matters of the *highest economic and social importance* aiming to ensure intergenerational finances" aimed at, amongst others, "[making] pensions at a time of great pressure on public finances, and [reflecting] changing demographics, life expectancy and social conditions".[124] They concern measures which "dealt with controversial matters of huge political weight and clearly fall within the macro-political field".[125] These observations flow from its earlier observation that the *Joint Council* does not demand a court to illustrate with precision all the factors in play. Similarly, in *R (Drexler) v Leicestershire County Council*,[126] dismissing the appellant's appeal, the court held that the respondent's home-to-school transport policies for pupils with special educational needs, which differentiated between pupils aged 5 to 15 (free transport) and pupils aged 16 to 18 (subsidised transport only), did not

[124] *Delve* (n 118) [92] (emphasis added).
[125] ibid [93].
[126] *R (Drexler) v Leicestershire County Council* [2020] EWCA Civ 502, [2020] ELR 399.

amount to a breach of ECHR14, read together with art.8 and Article 2 of Protocol no 1 to the ECHR. The court observed that in the alleged instances of discrimination involving public expenditure issues, following *Carson*, "very weighty reasons" would have to be given by the government to justify the alleged instance of discrimination.[127] Simultaneously, however, courts have to recognise the "relative institutional competence" of the executive or the legislature on the one hand and the courts on the other, in the context of matters of public expenditure; these matters, calling for "political judgment", mean that the decisions rendered or measures enacted must be "[afforded] appropriate weight and respect".[128]

In contrast to the straightforward exercises that *Delve* and *Joint Council* demonstrate, the Supreme Court's approach in *SC* represents a turn to a more nuanced and complex reasoning process, which is more conducive to the explication of the ILPV. Writing for a unanimous seven-member court, Lord Reed modified the court's previous adoption of the MWRF standard as applied in, amongst others, *Humphreys*, *SG*, and *DA*. The revisions "reflect the nuanced nature of the judgment which is required".[129] Although the position remains that the government's decision in social and/or economic matters will generally be respected unless it is MWRF, the judgment clarifies, importantly, that the intensity of scrutiny may be strengthened depending on the circumstances of each case.[130] When a suspect ground of discrimination is engaged, "very weighty reasons will *usually* have to be shown, and the intensity of review will *usually* be correspondingly high".[131] Instead of being fixated with the label of the standard adopted,[132] it is important to

> […] *avoid a mechanical approach* to these matters, based simply on the categorisation of the ground of the difference in treatment. A more flexible approach will give appropriate respect to the assessment of democratically accountable institutions, but will also take appropriate account of such other factors as may be relevant… the courts should generally be very slow to intervene in areas of social and economic policy such as housing and social security; but, *as a general rule*, differential treatment on grounds such as sex or race nevertheless require cogent justification.[133]

[127] ibid [55].
[128] ibid [56].
[129] *SC* (n 25) [158] (emphasis added); see also *R (Pantellerisco) v Secretary of State for Work and Pensions* [2021] EWCA Civ 1454 [58].
[130] *SC* (n 27).
[131] ibid (emphasis added).
[132] See also text to nn 25–27 above.
[133] *SC* (n 25) [159] (emphasis added).

Compared to the previous more straightforward approaches in the United Kingdom and the HKCFA's approaches in *QT* and *Leung*, *SC* provides greater room for tackling more delicately legal and factual nuances brought by the competing factors affecting adjudicative exercises in the third step. It directs a judge to flesh out more clearly how the relevant factors may be said to have an impact on the weight to be given to one another. This in turn enables the ILPV of judges and the making of any value judgments therein to be explicated in a clearer manner, manifesting the heuristic potential of the third step. The departure from fixation with a "precise definition" of MWRF, and direct engagement with the underlying question (that is, the scope of the margin of judgment), provides greater guidance on explicating what consideration is given to each relevant factor and their contribution towards the judge's reasoning process leading up to adopting a standard of review.

C. Towards a more delicate articulation of reasons

The foregoing comparative analysis between the approaches to the third step reveals a general shift from straightforward applications of legal standards to more delicate formulations. As a matter of a judge's legal analysis and communicating to the audience his reasoning process, such a move provides an opportunity for improving the heuristic and communicative potential of the third step. This is because, as Nicola Lacey observes, lawyers' inclination to "construct the world in terms of dichotomized categories" is less likely to correspond to common-sense understandings, or how the interested parties perceive the reasoning process in the first place.[134] The departure from clear-cut standards towards a more holistic and flexible operation of the "sliding scale" therefore enables judges to explicate to the public more clearly how the competing factors gravitating a court towards different standards of review are handled. Not only would this be helpful in improving the communicative potential of the proportionality framework,[135] it would also provide more solid legal ground and analytical support for the balancing exercise to be undertaken in the final step of a proportionality test.

Critics may complain that the paper's demands on the clarity of reasoning and communication are quibbles, for it is the application of the correct standard *per se* that is key. This is most cogently illustrated in Binnie J's concurring

[134] Nicola Lacey, 'The Jurisprudence of Discretion: Escaping the Legal Paradigm' in Hawkins (n 73) 363, 371.

[135] See text to n 10 above.

judgment in *Dunsmuir v New Brunswick*[136] on the need to avoid excessive "lawyerly arguments" in judicial review:

> 133. […] The disposition of a [judicial review] case may well turn on the choice of standard of review. If the litigants do take the plunge, they may find the court's attention focused not on their complaints, or the government's response, but on lengthy and arcane discussions of something they are told is the pragmatic and functional test. Every hour of a lawyer's preparation and court time devoted to unproductive 'lawyer's talk' poses a significant cost to the applicant.…the law of judicial review should be pruned of some of its unduly subtle, unproductive, or esoteric features.
>
> 145. […] While a measure of certainty is inherent in the subject matter and unavoidable in litigation (otherwise there wouldn't be any), we should at least…(ii) get the parties away from arguing about the tests and back to arguing about the substantive merits of their case.

The apex authorities in both Hong Kong and the United Kingdom (barring *SC*) may be seen as providing support for this scepticism. With regard to the interwovenness of the articulation and conferral of weight to each relevant factor and the reasoning process leading up to a judge's conclusion on the adoption of a standard of scrutiny, however, such criticism inappropriately ignores the contributive hermeneutic and communicative roles played by the ILPV. Binnie J's criticism about the potential over-complication of legal arguments in judicial review ought to be treated with caution in the present context. While it does suggest that counsel appearing before a court should not be too fixated in articulating the precise wording for the standard of review to be applied, it does not diminish the importance on the part of judges to explicate his conclusion as to the adoption of a particular standard of review and justifying it in light of: (a) the unique facts of the case; and (b) the relevant precedents. As will be demonstrated below, clear explications of the reasoning process in the third step may also have an impact on a court's reasoning process under the balancing exercise.

[136] *Dunsmuir v New Brunswick* [2008] 1 SCR 190.

V. Balancing Exercise: A Deceptively Simple Orchestration of the Competing Interests

As illustrated in Section II(B) above, the balancing exercises in socio-economic rights adjudication under both Hong Kong and English law are skewed. Despite the difference in the precise reasons behind these exercises, the commonality between the skewedness may be summarised as follows: despite the apparent adoption of a "balancing" exercise, the government's position is likely to be generally preferred—in that the aggrieved applicant would face a higher hurdle to satisfy than a traditional balancing test (in the sense that both ends of the balance are accorded equal weight). This in turn risks mystifying the heuristic potential of the balancing exercise, for the so called "balance" does not, in effect, afford equal consideration to both sides of it.

A. The impact of the skewed approaches in Hong Kong law and English law: same, same but different

The expressly skewed nature of the balancing exercises in Hong Kong law has resulted in a lack of clarity of expression on the competing rights and concerns involved. The two HKCFI judgments canvassed above, *Infinger* and *Ng Hon Lam Edgar*, exemplify this problem. In *Infinger*, Chow J concluded that, "for the same reason" that he employed to conclude that the impugned differential treatment is not justified, a fair balance has not been struck by the differential treatment under the policy in question, and hence the policy was unlawful.[137] While this conclusion on the facts of the case is correct (the government having failed to pass the third step), this reasoning is analytically problematic. The fact that the fourth step is where the infringed right is given full consideration vis-à-vis the legitimate aims of the policy means that it ought to be given independent articulation and elaboration as to its contribution to a judge's reasoning, instead of being subsumed under the third step.

In a similar but slightly different vein, Chow JA in *Ng Hon Lam Edgar* identified one factor on each side of the balance: (a) there would only be a "very limited increase in the number of HOS flats which may become additionally available to heterosexual couples to purchase as a result of the [Spousal Policy]"; and (b) the "unfair or unreasonableness" inflicted upon the applicant as a result of the policy in question.[138] He then immediately proceeded to conclude that the impugned policy operates on the aggrieved applicant "with such oppressive

[137] *Infinger* (n 95) [51(4)]–[51(5)].
[138] *Ng Hon Lam Edgar* (n 98) [76].

unfairness that it cannot be regarded as a proportionate means of achieving the [legitimate aim]".[139] While the notion of "oppressive unfairness" follows the language in *Hysan*,[140] the observation about the limited nature of achievements brought by the legitimate aim falls short of illustrating how it contributes to the value judgment made about the "oppressive" and/or "unfair" nature of the impugned treatment. The conclusion reached in this balancing exercise is therefore, with respect, not much different from a bare assertion of the cardinal importance of the right of equality based on the unfairness inflicted upon him, without illustrating as to how the other side of the balance (legitimate aims) contribute to this process of reasoning.

The extension of MWRF into the balancing exercise under English law in the context of socio-economic rights have produced problems that are similar to, but not as extensive as, those observable in *Infinger* and *Ng Hon Lam Edgar*. A key post-*DA* judgment where step four was considered in detail was *Joint Council*. This case concerned a scheme under the Immigration Act 2014, which imposed an obligation on landlords to take measures to provide private accommodation to tenants who were disqualified from obtaining so as a result of their immigration status. The government successfully challenged the lower court's decision the scheme was incompatible with ECHR14 (read with Article 8 of the ECHR). In adopting the MWRF standard at the balancing stage of the proportionality test and concluding that the impugned differential treatment is justified,[141] the court discussed in painstaking detail the relevant interests and why the individual rights engaged are not as important as others. After noting that "very considerable deference" should be given to Parliament's assessment of public interest and that the precise impact of the policy in question is difficult to quantify,[142] the court held that Parliament was aware of the risks of discrimination by landlords in implementing the scheme; it would, therefore, be improper to speculate what the Parliament might have expected.[143] After stating that "discrimination in all its forms is, of course, abhorrent", the court countenanced that the discrimination (and its risks) on the facts of the case emanate not from the policy itself but from the landlords, that is, private individuals' execution of the scheme.[144] In particular, the court also noted expressly at multiple junctures that the design and enforcement of the welfare scheme—a matter for Parliament—means that "great weight" should usually be given to it as the decision-maker.[145]

[139] ibid.
[140] *Hysan* (n 6) [78].
[141] *Joint Council* (n 115) [134].
[142] ibid [143]–[145].
[143] ibid [147].
[144] ibid [148(i)]–[148(ii)], [149(iii)], [150].
[145] ibid [146]–[147], [148(iv)], [149(iv)].

Compared to *Infinger* and *Ng Hon Lam Edgar*, *Joint Council* provides a fruitful example of how clear explications of the ILPV when articulating the weight to be attached to factors on each side of the balance, in light of the skewed focus, illuminate a judge's reasoning process. Such skewed focus of the balancing exercise highlights two matters. First, the apparent conflict between a "balance" and the skewed focus need to be addressed. Secondly, and more importantly, it is pertinent for a court to communicate to its audience how the infringed right remains being given adequate consideration in terms of its (alleged) importance vis-à-vis the legitimate aims pursued by the impugned measure. Conclusions reached in the third step (sliding scale) may be of assistance in characterising the nature of the legitimate objectives vis-à-vis the impugned right. But, to merely follow the conclusions reached in the third step (as in *Infinger*) would be an inappropriate simplification of the relations between the infringed right and legitimate aims pursued that a court is required to sketch under the fourth step.

B. Defending against skewness: the criticality of clear articulation of the relationship between competing forces of tension

The skewed focus of the balancing exercises sketched above ostensibly conflicts with the notion of a "balancing" exercise—in the sense that matters influencing the judge's decision about whether the impugned measure satisfies the proportionality test should be treated as, at least on their own, equal considerations. This, on its own, is not sufficient for a satisfactory articulation of the judge's ILPV in the reasoning process, for it is his explanation for: (a) why certain factors are seen as more important than others on the balancing scale; and (b) why, despite factors suggesting the contrary, the opposite conclusion is reached. Absent any changes to the formulations of the aforesaid balancing exercises, the ILPV would serve as a useful methodological injunction remedying the ostensibly disproportionate focus on and preference for maintaining the validity of the impugned measure.

The plurality of values that exist under this step means that, in assigning weight to each identified factor relevant to the balance, subjectivity on the part of a judge is inescapable.[146] Mere references to consequences and evaluative

[146] Indeed, both Dworkin's and MacCormick's account of adjudication, briefly canvassed above, acknowledge and envisage a version of value pluralism that there exists a multiplicity of values, some of which conflict with one another: Neil MacCormick, 'Contemporary Legal Philosophy; The Rediscovery of Practical Reason' (1983) 10 Journal of Law and Society 1, 13–14; Ronald Dworkin, *A Matter of Principle* (Clarendon Press 1985) 143–145. Traces of such plurality are evident in the choices open to a Judge to opt from in the third and fourth steps of a proportionality analysis, as analysed above.

considerations on the part of the judge are more likely than not to be sufficient. This is because they only provide general predictions as to what one might expect to find in the judge's reasoning process, but not the reason that the judge deems most compelling for justifying his adoption of a particular conclusion.[147] On the other hand, express acknowledgement of the legal value judgment involved—whilst unlikely to lead to an outcome that each litigating party would be satisfied with—would at least render the reasoning process a lot more transparent than, for example, a bare claim that the reasons adopted in the third step are equally applicable to the reasoning process under the balancing exercise. The diversity of views embodied in the various rights and interests relevant to the balancing exercise, whether in its skewed form or in a MWRF form, allows a judgment to be demonstrated as being "sensitive to the frictions and stresses of [law's] intellectual sources".[148] Not only does this help to illustrate that the balancing exercise does properly reflect the competing concerns in question (each being valid in their own ways), it also "fosters public discourse"[149] in clarifying the relationship between the infringed right and the objectives of the impugned measure.

VI. Conclusion

Diametrical oppositions of rights and interests in the proportionality analysis calls for a high degree of clarity when defending adjudicative choices . The comparative analysis of approaches in Hong Kong and English law above has demonstrated the potential confusion brought by a lack of sufficient revelation of one's ILPV. Although stronger protection for the rights of aggrieved individuals would certainly contribute towards more robust protection of human rights, this ought not to come at the expense of clarity in analysis and reasoning, for it would risk an impression on the part of the audiences of not affording sufficient consideration for countervailing factors that call for a different standard of review. Clearer spelling out of the process of judicial reasoning, as observable from *Infinger*, *Ng Hon Lam Edgar*, and *SC*, are commendable approaches towards ensuring that justice is not only done, but also seen to be done.

In light of the phenomena of "entirely associating the integrity of a legal system with the outcome…of cases determined by the courts",[150] it is all the more

[147] William Lucy, 'Adjudication for Pluralists' (1996) 16(3) Oxford Journal of Legal Studies 369, 383–385.

[148] Dworkin (n 67) 88–89.

[149] Barak (n 36) 463.

[150] Geoffrey Ma, 'Criticism of the Courts and Judges: Informed Criticism and Otherwise' (2018 Supreme Court of Queensland Oration, Queensland, 21 May 2018) § 24 <https://www.hkcfa.hk/en/documents/publications/speeches_articles/index.html> accessed 20 February 2022.; Baroness Chakrabarti CBE, 'Walking the Tightrope of Independence in a Constitutionally Illiterate World' in Jeremy Cooper (ed), *Being a Judge in the Modern World* (OUP

desirable for judges to, instead of shying away from such concerns behind a veil of following precedents *per se*, illustrate how they reason through the well-known elements of a proportionality test and the choices they have made therein. The ILPV serves as a practical and accessible tool for judges to demonstrate their critical awareness of reasoning through the choices that they make in adjudication and attempt to admit and defend the subjectivity therein at the same time. Its encouragement of frank admission of subjectivity and emphasis on the defensibility of legal choices made in the course of reasoning and adjudication serves to enhance the transparency of legal reasoning and explicate accessibly the very intelligibility of the law to the general public. That said, its inevitable downside is that it falls short of offering a panacea for individual or governmental dissatisfaction against adverse outcomes in proportionality assessments.

2016) 37, 42. An example of such a quick jump to conclusion may be found in the Government's hostile reaction to the declaration of incompatibility granted in *R (F) v Secretary of State for the Home Department* [2010] UKSC 17, (2011) 1 AC 331 against Part 2 of the Sexual Offences Act 2003 (i.e. the Sexual Offences Register).

Cambridge Law Review (2022) Vol VII, Issue 1, 33–56

Jumping into the SPAC Race: Protecting the UK Retail Investor

AKANKSHA MARIA PAUL*

ABSTRACT

The year 2020 witnessed special purpose acquisition companies (SPACs) emerge as the hottest investment offering on the global capital markets. The popularity of SPAC structures can be attributed to a combination of factors such as the pandemic induced low interest rate environment, downturn in the business cycle and growing number of equity starved companies. Though traditionally associated with 'pump-and-dump' schemes, the SPAC wave of 2020 was characterised by SPACs backed by optimistic projections and celebrity endorsements, which attracted the attention of 'mom-and-pop' retail investors. It is argued in this paper that while SPACs provide retail investors a relatively 'riskless' investment option in volatile market conditions, such unsophisticated investors do not always understand the risk and reward structure of SPACs and the dangers of optimistic projections. Consequently, concerns regarding retail investor interest protection have led to increase in regulatory scrutiny and securities litigation in jurisdictions such as the United States. Despite these concerns, on 27 July 2021, the UK's Financial Conduct Authority introduced certain changes to the Listing Rules in a bid to enter the SPAC race. The paper analyses UK's changing SPAC regulatory landscape through the lens of agency issues and concerns regarding retail investor protection.

Keywords: special purpose acquisition company; retail investor; principal-agency issues; initial public offers; financial regulation

* LLM (Corporate and Securities Law) (2020-21), London School of Economics and Political Science. I am grateful to Professor Simon Witney for his valuable comments on earlier drafts. Any errors that remain are my own. paulakankshamaria@gmail.com.

I. INTRODUCTION: THE SPAC WAVE OF 2020

The year 2020 saw special purpose acquisition companies (SPACs) emerge out of the shadows of the financial world[1] as the hottest offering on Wall Street.[2] SPACs dominated the US Initial Public Offering (IPO) landscape in 2020 with companies such as BuzzFeed Inc[3] and Nikola Motor Co.,[4] aiming to go public through the SPAC route. As of the first quarter of 2021, global IPO issuance was at $202.9bn, primarily fuelled by SPAC activity in the United States.[5] It should be noted that SPACs are not a product of ingenious financial engineering born out of the exigencies of the pandemic, but have been around since the 1980s.[6] Historically associated with 'pump and dump' schemes,[7] the SPAC boom of 2020 was characterised by investment vehicles backed by 'optimistic projections,' star-power[8] and popularity among retail investors.[9] Well known private equity and hedge fund operators along with celebrities such as Shaquille O'Neal,[10] Jay-Z,[11] Jennifer Lopez and Alex Rodriguez[12] are some of the popular names that backed SPAC structures. It is argued in this essay that 2020 saw a metamorphosis of SPACs into an

1 Miles Kruppa and Ortenca Aliaj, 'A Reckoning for SPACs: Will Regulators Deflate the Boom' *Financial Times* (New York, 4 May 2021) https://www.ft.com/content/99de2333-e53a-4084-8780-2ba9766c70b7 accessed on 25 July 2021.

2 Aaron Elstein, 'SPACK Attack' (2021) 37(6) Crain's New York Business <https://www.proquest.com/trade-journals/spac-attack/docview/2490739587/se-2?accountid=9630> accessed on 28 July 2021.

3 Benjamin Mullin, 'BuzzFeed Reaches Deal to Go Public Via SPAC, Acquire Complex Networks' *Wall Street Journal* (24 June 2021) <https://www-wsj-com.cdn.ampproject.org/c/s/www.wsj.com/amp/articles/buzzfeed-nears-deal-to-go-public-via-spac-eyeing-digital-media-rollup-11624485898> accessed on 25 July 2021.

4 Opinion Lex, 'SPACS/ Nikola: Fresh-baked Fruitcake' *Financial Times* (23 December 2020) <https://www.ft.com/content/60282388-3b8d-40b0-9b9a-130574678d93> accessed on 25 July 2021.

5 PwC, *IPO Watch Europe Q1 2021* <https://www.pwc.co.uk/audit-assurance/assets/pdf/ipo-watch-europe-q1-2021.pdf> accessed on 25 July 2021.

6 Usha Rodrigues and Mike Stegemoller, 'Exit, Voice, and Reputation: The Evolution of SPACS' (2013) 37 Delaware Journal of Corporate Law 849, 875.

7 ibid 875.

8 Kruppa (n 1).

9 Harriet Agnew and Ortenca Aliaj, 'SPAC Boom is Creating 'Castles in the Sky,' Jim Chanos Warns' *Financial Times* (London, 25 June 2021) <https://www.ft.com/content/da44b18e-51e5-40ab-9e34-70879952edce> accessed on 25 July 2021.

10 Brooke Masters, 'Year in a Word: SPACS' *Financial Times* (December 31, 2020) <https://www.ft.com/content/80458983-1693-4022-ba23-113925d24d70> accessed on 25 July 2021.

11 Due Diligence, 'SPACs vs Short Sellers: The Great Money Grab of *2021'Financial Times* (New York, 18 March 2021) <https://www.ft.com/content/c94f51f5-c042-42a6-8ba1-81b5672d2820> accessed on 25 July 2021.

12 Kruppa (n 1).

investment offering appealing to the public markets, leading to concerns regarding the protection of retail investors.

The 2020 surge in SPAC activity can be attributed to uncertain market conditions, reduced IPO activity, need for capital and a low-interest rate environment caused by the COVID-19 pandemic.[13] The pandemic-induced downturn in the business cycle has dulled the hope for equity-starved companies to access the public markets.[14] SPACs have stepped in to fill this gap, offering a fuss-free way to listing and accessing equity. For sponsors, SPACs offer an attractive vehicle to deploy funds, given the current climate of low interest rates and high market valuations.[15] Furthermore, market distress caused by the pandemic has made it easier to find and acquire viable targets. Finally, SPACs have stepped in to provide the SPAC investor an alternative investment source, allowing for public participation and co-investment alongside an experienced sponsor in a private equity style deal. From a retail investor perspective, SPACs claim to offer a relatively 'riskless' investment in volatile market conditions with tremendous upside potential due to high levels of liquidity.

These factors coupled with the need to re-assess traditional IPO processes and chart alternate paths for companies to access the capital market, make SPACs an attractive offering. The deal that 'put SPACs back on the map' was the 2019 Virgin Galactic's merger with Social Capital Hedosophia, a SPAC sponsored by the Facebook alum Chamath Palihapitiya.[16] The deal gave the investment vehicle a stamp of credibility and publicity among capital market participants and opened the floodgates for SPAC incorporations, particularly in the United States. Optimistic press reports highlighting few successful and highly visible celebrity-backed SPACs further fuelled this wave.[17]

The renewed interest in SPACs in the US markets has resonated globally leading to increased confidence in SPAC structures among sponsors, investors, and targets.[18] This has led to regulatory competition among jurisdictions to emerge as SPAC incorporation magnets. In Europe, Euronext Amsterdam,

[13] Thomas Vita, Fiona Millington and Kevin Connolly, 'SPACs: The London Alternative' (*Norton Rose Fulbright Publications*, October 2020) <https://www.nortonrosefulbright.com/en-gb/knowledge/publications/94734f5e/spacs-the-london-alternative#section3> accessed on 25 July 2021.

[14] Hugh Osmond, 'Time for UK Regulators to Open Door to SPACs' *Financial Times* (London, 17 December 2020) <https://www.ft.com/content/b364f03e-b026-4ec5-82fb-3991400de851> accessed on 27 July 2021.

[15] Vita (n 13).

[16] Elstein (n 2).

[17] Michael D Klausner and Michael Ohlrogge and Emily Ruan, 'A Sober Look at SPACs' (October 28, 2020). Yale Journal on Regulation, Forthcoming, Stanford Law and Economics Olin Working Paper No. 559, NYU Law and Economics Research Paper No. 20-48, European Corporate Governance Institute – Finance Working Paper No. 746/2021 <https://ssrn.com/abstract=3720919> accessed on 27 July 2021.

[18] PwC Report (n 5).

Frankfurt and Stockholm exchanges are emerging as hotspots for SPAC activity.[19] The UK, however, was slow to ride the SPAC wave. As of July 2021, there have been just over 50 SPACs listed on the London Stock Exchange (LSE), characterised by a small number of large listings.[20] The slow SPAC activity in the UK has been attributed to certain regulatory blockades in the Financial Conduct Authority (FCA) Listing Rules (Listing Rules)[21] that provide for trading of SPAC shares to be suspended once a target is identified.[22] This peculiar feature of the London market has made the UK SPAC unattractive to sponsors and investors.[23] Given, however, the increasing financial competitive edge of other European jurisdictions in a post-Brexit world, the UK Listing Review (Hill Review) was commissioned to propose revisions to the Listing Rules in a bid to transform London into an attractive listing venue for SPACs.[24]

The regulatory overhaul of the UK Listing Rules has, however, come during the ebbing of the SPAC wave on Wall Street, as regulatory authorities, sponsors and investors grow increasingly queasy about SPACs.[25] With a fall in total SPAC activity in the second quarter of 2021, many have questioned whether the SPAC wave was really just a bubble, bound to burst as interest rates improve.[26] Furthermore, there has been increasing concern regarding the protection of the interests of unsophisticated retail investors, who may be unaware of the risks and reward structure of SPACs and the dangers of optimistic projections.[27] The US has witnessed an increase in regulatory scrutiny of SPACs by the SEC[28] and increase in securities litigation by public shareholders alleging misstatements, fraud and breach of fiduciary duties by directors of companies that have gone public through the SPAC route.[29] Despite, however, the growing concern regarding retail investor protection, the UK is diving headfast into the SPAC race with the introduction of

[19] ibid.

[20] Financial Conduct Authority, *Investor Protection Measures for Special Purpose Acquisition Companies: Proposed Changes to the Listing Rules Consultation Paper,* April 2021 (CP 21/10) 12.

[21] FCA Listing Rules (July 2021) <https://www.handbook.fca.org.uk/handbook/LR.pdf> accessed on 1 August 2021.

[22] FCA Consultation Paper (n 20).

[23] ibid.

[24] Baker McKenzie, *H1 IPO Snapshot: Unfolding Trends for 2021,* <https://www.bakermckenzie.com/-/media/files/insight/publications/2021/06/ipo-h1-snapshot-2021_final.pdf?la=en> accessed on 23 July 2021.

[25] ibid 6.

[26] Ivana Naumovoska, 'The SPAC Bubble is About to Brust' (*Harvard Business Review,* 18 February 2021) <https://hbr.org/2021/02/the-spac-bubble-is-about-to-burst> accessed on 17 July 2021.

[27] Kruppa (n 1).

[28] Brooke Masters, 'Grab is Lunging for the Top of the SPAC Market' *Financial Times* (14 April 2021). <https://www.ft.com/content/df14b6f5-a967-4f0f-8441-a5ab48d66dec> accessed on 31 July 2021.

[29] Kruppa (n 1).

the FCA Policy Statement PS21/10 (Policy Statement) on 27 July, 2021, containing the final changes to the Listing Rules (Revised Listing Rules).[30]

Amid the growing SPAC buzz and revision of the Listing Rules, this paper attempts to re-focus attention to the agency issues in a SPAC structure and the concerns regarding the protection of retail investors. Section I of this paper analyses the concept and structure of a SPAC. Section III aims to identify the agency issues in a SPAC and explore the concerns regarding retail investor protection. Section IV analyses UK's changing SPAC regulatory landscape and whether the same adequately protects the interests of SPAC retail investors.

This paper aims to extend the literature on the agency issues in a SPAC. As far as the author is aware, there has been no comprehensive analysis of the concerns and protection of retail investors in the context of the SPAC wave of 2020 and the Revised Listing Rules.

II. The SPAC and the Retail Investor

SPACs have a notorious ancestry and are associated with blank cheque companies of the 1980s, which were used to defraud unsophisticated investors in the United States.[31] The US Securities Enforcement Remedies and the Penny Stock Reform Act 1990, along with strict regulation of blank cheque companies, lead to a decline of this investment form in the 1980s.[32] It was not until 1992 when the modern SPAC structure, with its various in-built investor protection mechanisms, was formulated in a bid to gain the approval of the US Security and Exchange Commission (SEC).[33] During the mid-2000s, a decline in traditional IPO activity in the United States, saw an increase in the use of the SPAC form.[34] Meanwhile, the UK witnessed a surge in SPAC activity after the financial crisis of 2008.[35]

In the UK, SPACs are considered as cash shell companies which do not meet certain independence and operating history requirements for premium listing. UK SPACs are listed mainly on the Standard segment of the Official List or on the Alternate Investment Market (AIM) of the LSE.[36] Given, however, certain

[30] Listing Rules (Special Purpose Acquisition Companies) Instrument 2021 (FCA 2021/29), Annex B <https://www.handbook.fca.org.uk/instrument/2021/FCA_2021_29.pdf> accessed on 1 August 2021.

[31] Derek K. Heyman, 'From Black Check to SPAC: The Regulators' Response to the Market and the Market's Response to the Regulation' (2007) Entrepreneurial Business Law Journal 531, 532; Rodrigues (n 6) 875.

[32] ibid 532; Rodrigues (n 6) 875.

[33] Derek (n 31) 540.

[34] ibid 532.

[35] Winston & Stawn LLP, 'SPAC to the Future: The Recent Resurgence of UK SPACS and Latest Trends' (2018) < https://www.winston.com/images/content/1/4/v3/142888/SPAC-to-the-Future-the-Recent-Resurgence-of-UK-SPACs-and-Latest.pdf> accessed on 31 July 2021.

[36] ibid.

requirements of the AIM regime such as shareholder approval for reverse takeovers, the Standard segment is the most popular option for UK SPACs.[37]

A. SPAC STRUCTURE

In the US, SPACs are designed to track the protective requirements of Rule 419 of the Securities Act, 1933. SPACs in most jurisdictions borrow heavily from the US SPAC structure with certain deviations. A typical SPAC is a blank-check or clean cash shell company with no assets or operating history.[38] A SPAC is founded by sponsors and taken public with the sole aim of acquiring a yet-to-be-identified target company (mostly private companies) within a short time-horizon,[39] usually twenty-four months. The target is then acquired by the SPAC and taken public through a reverse merger (De-SPAC Transaction).[40] Unlike in a typical IPO, the initial SPAC IPO is swift and relatively cheap due to the lack of lengthy disclosures and compliance with other listing requirements.[41]

In a SPAC IPO, a SPAC investor is usually offered a 'unit' that is, a combination of shares and share warrants.[42] The issue proceeds are usually held on trust in an escrow account.[43] Warrants are granted to enable SPAC investors a right to acquire additional shares of the post-acquisition company, at a specified time in the future, at a pre-agreed strike price, which is usually a 15% mark-up of the IPO share issue price.[44] Similar to traditional private equity, to ensure that SPAC sponsors and managers (hereinafter collectively, SPAC Sponsors or Sponsors) have adequate 'skin in the game,' Sponsors initially buy-in a percentage of the SPAC preferred shares (usually 20% of the post IPO equity) (Founder Shares) at a low valuation which is also then placed in escrow.[45] During the SPAC IPO, the Sponsors may be further issued a combination of ordinary, Founder Shares and warrants which is subject to a lockup period.[46]

Upon completion of a successful acquisition, Sponsors may end up owning 20% of the newly acquired company.[47] If the acquisition mandate, however, is not

37 Vita (n 13).
38 Rodrigues (n 6) 871.
39 ibid.
40 Rodrigues (n 6) 871.
41 ibid.
42 Magnus Blomkvist and Milos Vulanovic, 'SPAC IPO Waves' (2021) Economics Letters 197; Rodrigues (n 6) 871.
43 Rodrigues (n 6) 871.
44 Vita (n 13).
45 Rodrigues (n 6) 871.
46 Vita (n 13).
47 ibid.

met in the given timeframe, Sponsors forgo their shares. [48] In addition to the above, SPAC Sponsors and certain other sophisticated SPAC investors may also purchase additional shares and warrants through private placement, which supplements the amount of their "skin in the game"[49] and helps flush the SPAC with additional capital, if required.

It must be mentioned that there are significant structural differences between a US and a UK SPAC. Firstly, with respect to the investor approval mechanism once a target is identified, in the US a SPAC shareholder may vote to either approve the acquisition or vote against it. If approved, the shareholder comes to hold shares in the acquired company, which then trades publicly through the reverse merger.[50] If the acquisition is not approved, the shareholders may elect to redeem their investment and are returned pro rata their share of funds in the escrow account. Given that the De-SPAC Transaction heralds the end of the SPAC, SPACs may often bargain for a positive vote on a proposed acquisition in several ways. For example, by promising to repurchase the shares held by the SPAC investor upon competition of the De-SPAC Transaction[51] or by an open market purchase of public shares by SPAC Sponsors.[52] Whereas for a UK SPAC listed on the Standard segment, there is no requirement of seeking shareholder approval for acquisitions. Instead, post-acquisition, there is a requirement for a prospectus of the combined entity to be re-filed for FCA approval.[53]

Secondly, under the Listing Rules, SPAC shares are suspended from trading from the date of announcement of the De-SPAC Transaction till the re-filing of the prospectus of the combined entity.[54] This means that once a potential target is identified, Investors in a UK SPAC are locked into their investments. Thirdly, UK SPAC investors are typically not granted redemption rights.[55] In the United States, under the NASDAQ and NYSE rules, those SPAC shareholders who vote against the acquisition proposal must be given an option for redeeming their shares.[56] From a US investor perspective, this redemption right makes SPACs a cheap and liquid investment. Fourthly, in the US, there is a strict investment criterion for the SPAC IPO proceeds.[57] These proceeds are invested in US treasury

48 Rodrigues (n 6) 871.
49 ibid 895.
50 Rodrigues (n 6) 871.
51 ibid 872.
52 Tim Jenkinson and Miguel Sousa, 'Why SPAC Investors Should Listen to the Market' (2011) Journal of Applied Finance 38, 50.
53 Listing Rules (n 21), LR 5.6.
54 Listing Rules (n 21), LR 5.6.4R, LR 5.6.5A R and LR 5.6.5G.
55 Thomas Vita (n 13).
56 ibid.
57 Ramey Layne and Brenda Lenahan, Vinson & Elkins LLP, 'Special Purpose Acquisition Companies: An Introduction' *Harvard Law School Forum on Corporate Governance* (6 July 2018)

bonds and earn interest. By contrast, the UK has no such investment requirement for the deployment of IPO proceeds and therefore offer significant flexibility on the short-term use of such proceeds. As will be explained in Section IV of this paper, these features of the UK SPAC presented additional challenges in ensuring that the interests of SPAC retail investors are protected and have subsequently been dealt with under the Revised Listing Rules.

The design of a SPAC vehicle enables it to serve the interest of a wide variety of market participants. Rodriguez and Stegemoller (2013) characterise SPACs as a 'poor man's private equity fund,' that offers a chance to 'mom-and-pop' investors to finance a management team's hunt for a target.[58] From the perspective of the SPAC Sponsor, SPACs offer a cheap method to access and raise money from the capital markets to finance their quest for a target.[59] Finally, through the De-SPAC Transaction, SPACs provide the management of a small private company retention of control while infusing cash and opening a back-door to the capital markets.[60]

B. THE RETAIL SPAC INVESTOR

The SPAC wave of 2020 witnessed a surge in SPAC listings and increased participation by retail investors. In the US, as of February 2021, retail investors accounted for 40% of all trading in SPACs.[61] According, however, to Mitchell Littman, a partner at a New York law firm, SPACs are only meant for certain types of investors. "This is not something that anybody should be putting their 401(k) or (individual retirement account) into," said Littman.[62] It is therefore important to explore the peculiar characteristics of SPAC retail investors that warrant special regulatory protection.

SPAC retail investors can be categorised into (a) mass retail investors; (b) mass affluent investors; (c) high net worth individual investors; and (d) ultra-high net worth individual investors.[63] While (c) and (d) are usually backed by

≤https://corpgov.law.harvard.edu/2018/07/06/special-purpose-acquisition-companies-an-introduction/#2b> accessed on 1 August 2021.

58 Rodrigues (n 6) 874.

59 ibid.

60 ibid.

61 Ortenca Aliaj and James Fontanella Khan, 'Retail Investor Apathy Threatens to Derail SPAC Deals' *Financial Times* (New York, 10 March 2021) <https://www.ft.com/content/7554fead-6784-421a-8659-79afc8fbeeed> accessed on 17 July 2021.

62 Scott Malone, 'RPT-IPO View – Crunch Time Coming for Blank-Check Companies' *Reuters* (26 March 2006) <https://www.littmankrooks.com/pdf/Crunchtime-for-blank%20checkIPO.pdf> accessed on 31 July 2021.

63 Jonathan V. Beaverstock, Sarah Hall and Thomas Wainwright, 'Scoping the Private Wealth Management of the High Net Worth and Mass Affluent Markets in the United Kingdom's Financial Services Industry' (May 2010) <https://www.nottingham.ac.uk/business/who-we-are/centres-and-institutes/gcbfi/documents/researchreports/paper71.pdf> accessed on 31 July 2021.

sophisticated private wealth management teams, the concern of this paper is the 'mom and pop' retail investor, which lacks both financial sophistication and a large amount of capital at its disposal. Often such retail investors jump from one hot offering to another. For example, in 2020, Wall Street saw retail investors jump between big technology stocks, sustainable investments and SPACs.[64] Increased retail participation in SPACs has been attributed to this 'bandwagon effect' and a 'fear of missing out' on the action. Retail investors lack sophistication to understand the lifecycle, incentive and reward structures of SPACs and are often the risk bearers in a bad SPAC deal. These issues will be discussed further in Section III.

Irrespective of the jurisdiction, concerns regarding retail investor protection in SPACs arise due to two main issues that is, collective action problem and the issue of information asymmetry. Firstly, like in the case of a company with dispersed shareholding, retail investors in a SPAC suffer from the collective action problem. This impacts the SPAC and investor dynamics in several ways. For example, in the US, where proposed acquisitions require shareholder approval, SPAC Sponsors face logistical issues in solicitation and engagement with retail shareholders, slowing down deals and extensions, thus affecting internal SPAC dynamics.[65] Secondly, opaque SPAC structures and reduced regulatory oversight exacerbate issues of information asymmetry which have given rise to several instances of litigation regarding securities in the US. For example, a petition was filed against Nikola Corp, the poster child of the SPAC 2020 boom, for making false and misleading statements to its retail investors.[66]

Considering the above issues, the US House of Representatives Subcommittee on Investor Protection, Entrepreneurship and Capital Markets agreed, on 24 May, 2021, on the need for SPACs to be regulated in the same way as traditional IPOs for the sake of protecting retail investors.[67] While regulators have become more sensitive to the protection of retail investors, there is a theory that such hyperactivity among retail investors is temporary and driven by the CoVID-19 pandemic. If so, once the ebb of retail investors crash, regulators need to ensure that investor protection measures put in place do not act as an effective barrier for

[64] Katie Martin and Robin Wigglesworth, 'Rise of the Retail Army: The Amateur Traders Transforming Markets' *Financial Times* (London, 9 March 2021) <https://www.ft.com/content/7a91e3ea-b9ec-4611-9a03-a8dd3b8bddb5> accessed on 17 July 2021.

[65] Aliaj (n 61).

[66] Claire Bushey and Ortenca Aliaj, 'Nikola Found Trevor Milton Charged with Making False Statements' *Financial Times* (Chicago, 29 July 2021) <https://www.ft.com/content/7469bb92-10c7-49a5-8646-193fb3777d2f> accessed on 17 July 2021.

[67] Joseph Williams, 'House Subcommittee Agrees More Investor Protections Necessary or SPAC Model ' *SNL Financial Services Daily* (25 May 2021) < https://www.proquest.com/trade-journals/house-subcommittee-agrees-more-investor/docview/2532412044/se-2?accountid=9630> accessed on 31 July 2021.

companies trying to access the capital markets through SPACs. Accordingly, keeping in mind this concern, Section IV of this paper will discuss the light touch investor protection regime proposed by the FCA in regulating SPACs in the UK and whether the same adequately protects the retail investor.

III. Retail Investors and Agency Cost Issues in a UK SPAC

The acquisition of a private target company by a SPAC is a form of a reverse merger.[68] In a traditional reverse merger, a successful private company is merged with a listed shell company, which is often a public virgin company incorporated for the sole purpose of the combination or a remnant of a previous operational public company.[69] In the case of SPACs, the cash-shell company is incorporated, flushed with public money with the sole objective of hunting private targets to take public. With no operating history or assets on which to base the investment decision, retail SPAC investors effectively purchase a management team and their ability to identify and purchase an acquisition target. Blomkvist and Vulanovic (2020) observe that this opaqueness is compounded by the 'one-shot-deal' nature of the SPAC entity, in which reputational factors of the SPAC Sponsor is no longer a tool to reel in agency costs and ensure investor protection.[70] Therefore, a SPAC is a blind pool of funds in which the SPAC investor does not know exactly what she is paying for.[71]

While the 2020 SPAC wave, backed by reputable sponsors and celebrity endorsements, solved the issue of reputation to a large extent, it is argued that even with an experienced management team, the structure of SPACs and nature of management incentives makes them replete with agency issues. Furthermore, as explained in Section II, it is argued that certain features of a UK SPAC such as (a) no investment guidelines on the proceeds of SPAC IPOs; (b) no shareholder approval process for proposed mergers; and (c) no redemption and weak exit rights for disapproving shareholders; exacerbate the issues of agency cost and protection of retail SPAC investors. This section will examine the misaligned interests of SPAC Sponsors and the retail SPAC investor and the consequent agency cost issues which require regulatory attention.

68 Naumovoska (n 26).
69 ibid.
70 Rodrigues (n 6).
71 Elstein (n 2).

A. SPAC IPO PROCEEDS AND THE PRPOBLEM OF FREE CASH FLOW

SPAC IPO proceeds constitute a source of free cash flow to the SPAC Sponsors, which are then utilised to hunt and acquire a target company. Jensen (1986) defines free cash flow as 'cash flow in excess of that required to fund all projects that have positive net present values when discounted at the relevant cost of capital'.[72] Jensen states that there exist managerial agency costs in ensuring that managers are motivated to utilise free cash flow responsibly, rather than wasting or investing the same below the cost of capital.[73] Developing on Jensen's idea of managerial agency costs relating to free cash flow, Schultz (1992) highlights that after a freshly-completed IPO, very strong incentives exist for managers to invest in negative net present value projects.[74]

To solve the issues of agency costs of free cash flow, SPAC issue proceeds are usually held in escrow. Furthermore, as per the SPAC prospectus, issue proceeds can be used only for certain purposes such as (a) acquisition of a company; (b) capital contribution to the merged company; (c) distribution in case of liquidation of the SPAC; or (d) redemption of shares.[75] Free cash flow can also lead to an issue of dilution for the SPAC investor if the SPAC utilises IPO issue proceeds for financing the operations of the acquired company or for redemption of shares.[76] To deal with this, SPAC Sponsors may often purchase additional SPAC shares and warrants through a pre-IPO private placement. This allows SPAC Sponsors to provide investors comfort that 100% of the issue proceeds will be kept in escrow and used to funding the acquisition, while the private placement proceeds will be used to finance the SPAC's operating expenses or fund the redemption of shares.[77] Additional funds to meet working capital arrangements may also be raised through Private Investment in Public Equity (PIPE) financing arrangements[78] or by disclosure in the prospectus, that IPO proceeds up to a certain percentage will be utilised for the purposes of working capital requirements.[79]

[72] Michael C. Jensen, 'Agency Costs of Free Cash Flow, Corporate Finance, and Takeovers' (1986) 76 The American Economic Review 323, 323.

[73] ibid 323.

[74] Paul Schultz, 'Unit Initial Public Offerings: A Form of Staged Financing' (1993) 34 Journal of Financial Economics 199, 200.

[75] Klausner (n 17).

[76] Rodriguez (n 6) 873 and 925.

[77] Mark Bonenfant, 'Special Purpose Acquisition Companies' *Buchalter Nemer* <https://www.buchalter.com/wp-content/uploads/2007/12/Special-Purpose-Acquisition-Companies.Bonenfant.pdf> accessed on 19 July 2021.

[78] Klausner (n 17) 49.

[79] FCA Consultation Paper (n 20) 17.

Agency costs created by the problem of free cash flow is further attempted to be reeled in through the provision of unit offerings in a SPAC IPO. As mentioned earlier, the SPAC IPO typically consists of an offering of shares and warrants. Schultz (1992) highlights that share warrants are usually used as a 'sweetener,' to make an issue more attractive by incentivising investors to subscribe into an IPO which otherwise maybe unattractive.[80] In the case of SPACs, share warrants are used to incentivise SPAC investors with the promise of shares in the post-acquisition company at the warrant's exercise price. It is argued that share warrants have a disciplining effect on the management.[81] Given that the purchase price of shares of the acquired company is at the pre-determined warrant exercise price (which is often above the market value of the shares), SPAC Sponsors are deterred from squandering IPO proceeds under the pretext of 'finding an appropriate target,' thereby preventing dilution of shareholder value.

The more funds held in escrow and shielded from the SPAC Sponsors, the more attractive and safer the SPAC is for the retail investor. Escrowing of issue proceeds to a large extent solves the issues of free cash flow and helps reduce investment risks.[82] SPACs may also obtain guarantees from banks and waivers from vendors, or the target from any claim on the amounts held in escrow as a further measure to protect trust value and the interest of retail SPAC investors.[83]

B. Problem of founder shares and time bound acquisitions

As stated in Section II of this paper, once listed a SPAC has, subject to extensions, twenty-four months to find and acquire a target company. The SPAC structure is formulated such that it must work against the clock to complete an acquisition or else face liquidation.[84] Rodrigues and Stegemoller (2013) argue that time constraints associated with SPACs have a disciplining effect on SPAC Sponsors, as investor money is not held indefinitely by the management in escrow, thereby preventing a decline in the value of the trust fund.[85] It is argued, however, that such time constraints create an artificial pressure on the SPAC Sponsor to complete an acquisition within a short time-frame, irrespective of its merits.

SPAC Sponsors receive 20% interest in the SPAC, which becomes valuable only if an acquisition is completed. Given the escrowing of Founder Shares and the limited time period within which the acquisition must be completed, SPAC

[80] Schultz (n 74) 200.
[81] ibid 201.
[82] Rodriguez (n 6) 914.
[83] ibid 913.
[84] Eric J Savitz, 'The New Blind Pools' *Barron's* (12 December 2005) 21.
[85] Rodrigues (n 6).

Sponsors have strong financial incentives to push a deal through, irrespective of whether it is the optimal choice.[86] Furthermore, unlike in traditional private equity where managers are rewarded upon the realisation of profit, in case of SPACs, Sponsors enjoy a large payday upon completing the acquisition.[87] Managers may even be offered employment contracts contingent upon completion of the acquisition mandate.[88] Therefore, under the typical SPAC structure, Sponsor rewards are almost completely divorced from the profitability of the acquired target once it is listed and hinges almost entirely on completing an acquisition within the given time frame. It is noted that while the average SPAC stock sank by 36% post-merger, SPAC Sponsors were able to generate an average return of about 400%,[89] representing the severe misalignment of interests between the SPAC Sponsors and investors while selecting a suitable SPAC target. Given that there is no formal requirement for shareholder approval in an UK SPAC, the chances of SPACs Sponsors pushing a bad deal through are significantly higher.

It has been argued that co-investment by Sponsors through pre-IPO private placement solves this issue of SPACs pushing a bad deal.[90] Given, however, that Sponsor contribution to the fund is an average of 2.5% of the IPO proceeds, the motivation of realising their 20% stake works as a stronger incentive to see an acquisition deal go through. Unless the Sponsor's ability to realise their 20% share upon acquisition is not contractually reduced or delayed, the same continues to present an agency cost issue. Alternatively, a legal requirement for detailed due diligence of the target or of a mandatory percentage of institutional shareholder participation may solve the issue of acquisition of bad targets. The later of the two approaches has been adopted by the Revised Listing Rules, as discussed in Section IV of this paper.

Finally, given the large number of SPACs which have been incorporated in 2020 and are seeking to fulfil their acquisition mandates in the next few months, the hunt for quality target companies is fierce.[91] This competition among acquirers make it harder to find healthy targets to acquire at fair valuation.[92] Savitz (2005) argues that potential targets realise that SPACs have a time-bound investment mandate which is often used against SPACs in negotiations regarding valuation, leading to issues of over pricing of unhealthy targets.[93]

[86] Derek (n 31) 549.
[87] Rodrigues (n 6) 894.
[88] Bonenfant (n 77).
[89] Elstein (n 2).
[90] Rodrigues (n 6) 896.
[91] Lulu Yllun Chen, 'SPACs See Asia as Next Hunting Ground for Takeover Targets' *Bloomberg* (31 January 2021) <https://www.bloomberg.com/news/articles/2021-01-31/spac-spending-spree-moves-to-asia-as-funds-seek-out-new-targets> accessed on 27 July 2021.
[92] Derek (n 31) 551.
[93] Savitz (n 84) 22.

C. Issue of information asymmetry

The problem of information asymmetry can arise in various stages of a SPAC's lifecycle. This paper identifies two such instances. Firstly, at the point of the SPAC IPO and secondly at the point of acquiring a target company. As mentioned in Section II, at the point of the SPAC IPO, a SPAC investor effectively buys into a management's ability to identify and acquire a target company. Unlike in a traditional IPO, a SPAC has no previous operating history or financials to disclose, lowering the level of information available and consequently increasing information asymmetries for the SPAC investor. It is argued that these information asymmetries are compounded for the retail investor as she is exposed to optimistic press reports, projections and lacks the sophistication or ability to scrutinise the same.

The traditional view of SPACs is that they are advantageous to companies who wish to go public but may find it difficult to convey information to the market or have difficulty in terms of valuing.[94] In this sense, SPACs have a unique role in bringing companies which have unusual businesses or few market comparables to the public markets, thereby addressing the issue of information asymmetry between these companies and the public markets. It is argued, however, that in the absence of any legal requirement of conducting high standard due diligence on potential targets, regulatory intervention or shareholder approval process for proposed acquisitions, the SPAC presents a problem of grave information asymmetry for the retail investor. It is argued that this is further compounded if retail investors have weak voting and exit rights (discussed below). As noted above, SPAC Sponsors have more incentive to push a bad deal through than do no deal at all. To minimise such information asymmetries, SPAC investors should be given sufficient informational rights at various stages of the SPAC life-cycle.

IV. Revamping UK's SPAC Regulatory Landscape and the Protection of Retail Investors

A. Background

The Listing Rules applicable to SPACs provide for a rebuttable presumption of suspension of trading of shares when a reverse takeover is in contemplation.[95] LR 5.6.7G provides that such situations will include where (i) the SPAC has approached a target's board; (ii) the SPAC has entered an exclusivity

[94] Klausner (n 17) 5.
[95] Listing Rules (n 21), LR 5.6.8G.

period with a target; or (iii) where the SPAC has given a target access to begin due diligence. The FCA states that in such situations, suspension of trading is necessary as the SPAC is unable to accurately assess its financial position and inform the market accordingly.[96] The suspension of trading serves to protect public investors from insufficient publicly available information which may harm market integrity and lead to opportunism and issues of insider trading.[97] The Listing Rules provide that such suspension can be avoided if sufficient market disclosures are made available following approval from the FCA.[98] The suspension period, however, can be long, effectively locking in dissenting shareholders till the deal is finalised and a revised prospectus is filed. The presumption of suspension, requirement of FCA approval and additional disclosures have been identified as regulatory blocks for SPAC listings in the UK.

Accordingly, as part of the Treasury's plan to strengthen UK's global financial position, the Hill Review chaired by Jonathan Hill was constituted to *inter alia* "improve the environment for companies to go public in London."[99] The Hill Review submitted its report on 3 March, 2021 and recognised that the UK had fallen behind the SPAC race, and required stronger public markets and an influx of growth company listings.[100] Accordingly, Recommendation 6 of the Hill Review states,

> Revise the Listing Rules which can require trading to be suspended in the shares of SPACs on announcement of a potential acquisition. Provide additional protections for shareholders at the time of the acquisition, such as a shareholder vote and redemption rights.

In light of the above recommendation, the FCA published its Consultation Paper CP21/10 in April 2021 (Consultation Paper). The Consultation Paper prescribed an "alternate route to the market for SPACs demonstrating higher levels of investor protection mechanisms."[101] The proposals were drafted by observing global market developments during the SPAC wave of 2020, recommendations by the Hill Review and stakeholder feedback.[102] The recommendations in the

96 FCA Consultation Paper (n 20) 26.

97 Financial Conduct Authority Technical Note, *Cash shells and special purpose acquisition companies,* (January 2018) <https://www.fca.org.uk/publication/ukla/tn-420-2.pdf> accessed on 19 February 2022.

98 Listing Rules (n 21) LR 5.6.8G.

99 UK Listing Review (3 March 2021) 11 <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/966133/UK_Listing_Review_3_March.pdf> accessed on 31 July 2021.

100 ibid 1.

101 FCA Consultation Paper (n 20) 4.

102 ibid 27.

Consultation Paper were aimed to remove barriers for listing of large SPACs and provide greater flexibility to SPACs which are backed by experienced management and have potential to reach a certain scale. Based on the responses received on the Consultation Paper, the FCA published its Policy Statement containing the Revised Listing Rules. The annexure to the Policy Statement contains the revised Primary Market Technical Note and the Listing Rules (Special Purpose Acquisition Companies) Instruments 2021 which amends the Listing Rules. These come into effect on 10 August, 2021. Considering the agency issues identified in Section III, this section will critically analyse whether the Revised Listing Rules properly protect the interest of retail SPAC investors in the UK.

B. Analysis of the Revised Listing Rules

(i) Exemption from the Presumption of Suspension of Trading

There exist strong incentives for SPAC Sponsors to push a deal through, irrespective of its merits, representing the misalignment of interest between the SPAC and its unsophisticated retail investors. As explained in Section III of this paper, these agency issues are exacerbated by a time-bound acquisition mandate and limited number of quality targets. It is argued that in such a climate, public investors need powerful signals to access the merits of a proposed acquisition.

The current regulatory regime for UK SPACs casts a presumption of suspension of trading of SPAC shares which locks in unhappy investors until a deal is closed. Currently, 40% of listed UK SPACs have their shares suspended.[103] It is argued that firstly, the lock-in deprives the retail SPAC investor of the advantages of high liquidity of the capital markets and imposes high opportunity costs on the retail investor who, could have otherwise withdrawn her investment from the SPAC. Secondly, SPAC investors are left in a period of uncertainty with limited or no information between the period of announcement and completion of the acquisition transaction. Thirdly, the lock-in has made the UK SPAC highly unattractive viz. other SPACs incorporated in US or Europe, which provide greater visibility, control, and liquidity to the public retail investor. Considering the points raised, it is argued that exemption from presumption of trading for SPACs with a higher degree of investor protection mechanisms, provides a valuable exit right for an unhappy retail investor and helps in the better allocation of capital in the public markets.

The Revised Listing Rules provide for a removal of the presumption of suspension of shares and thereby allows retail investors to react to the market.[104]

[103] FCA Consultation Paper (n 20) 26.
[104] Revised Listing Rules (n 30), LR 5.6.8(2).

Jenkinson and Sousa (2011) argue that the market can step in to provide signals to retail investors as to the exercise of their voting and exit rights between the period when a potential acquisition in announced and the decision date. Given that retail investors are unsophisticated, these market signals are valuable to solve the issue of information asymmetry and collective action. In an empirical study of US SPACs, it was observed that the SPAC share price between the acquisition announcement date and the decision date reflects the investors' assessment of the deal being proposed.[105] As on the decision date the SPAC share price reflects the market's evaluation of the proposed deal.[106] Subsequently, if the SPAC share is valued equal or below the trust value per share, the same is an indication of market signals that the deal is value destroying for the public shareholders. Conversely, a higher share price would indicate value creation. It is argued that such market signalling provides a good basis for retail investors to approve or reject a deal. It should be noted, however, that there are various methods in which SPACs can be structured to ensure that the deal is pushed through, irrespective of its merits. Furthermore, SPAC Sponsors may engage in vote buying from large institutional investors (who may be oppose to the proposed acquisition), closer to the approval date, in a bid to turn the vote positive. This would lead to an artificial demand in the SPAC's shares, thereby undermining the use of market signalling in the protection of retail investors.

(ii) Minimum Size Requirement

The Revised Listing Rules set a minimum size requirement for the aggregate gross cash proceeds raised in a SPAC IPO if it wishes to take the alternative route to suspension of trading. This has been set at £100m or more.[107] The purpose of the threshold is to judge a SPAC's ability to raise capital from large public investors. The FCA suggests that meeting such a threshold will require SPACs to have a high degree of institutional investment. The assumption being that institutional investors will (a) exercise a high degree of diligence when investing into a SPAC, without solely relying on inflated projections and celebrity optimism; and (b) heavily scrutinise potential acquisition proposals. The calculation of the threshold excludes any funds that SPAC Sponsors would have invested.

It is argued that an increased presence of institutional investors will lead to a higher level of due diligence at the time of the SPAC IPO and the De-SPAC Transaction. Presence of such sophisticated investors would lead to greater

[105] Jenkinson and Sousa (n 52) 41.
[106] ibid 42.
[107] Revised Listing Rules (n 30), LR 5.6.18A (1).

scrutiny of the investment proposals and provide adequate protection to the interests of retail investors. For example, the Financial Reporting Council's UK Stewardship Code, 2020 (Stewardship Code), lays down the framework for increased shareholder engagement and stewardship by large institutional investors. The Stewardship Code works on a 'comply or explain' basis where asset owners and managers must actively engage with issuers, holding them accountable on material issues, exercising rights and responsibilities with a view to create long-term value and sustainable benefits for the economy, environment, and society.[108] It is argued that an increased presence of public institutional investors in UK SPACs will help alleviate the issues faced by unsophisticated retail SPAC investors. As stated, however, in Section IV.B.(i), the role of large institutional investors in protecting retail investor interests must not be overstated. Moreover, the size of a SPAC is not relevant to the quality of its internal investor protection mechanisms or experience of its Sponsors.[109]

(iii) Ring-Fencing of SPAC IPO Proceeds, Redemption, and Repayment Process

The Revised Listing Rules provide that SPAC IPO proceeds should be 'adequately' ring fenced via an independent third party.[110] Furthermore, the purposes for which such ring-fenced funds can be used for are (a) funding an acquisition; (b) share redemptions; and (c) repayment of capital to public investors in case of SPAC liquidation or failure to meet the acquisition mandate. The Revised Listing Rules clarify that SPAC IPO proceeds may be used to fund working capital requirements, subject to adequate disclosure of specified amounts for such purpose in the prospectus filed at the time of the SPAC IPO.[111]

As highlighted in Section III of this paper, escrowing of SPAC IPO proceeds is essential to protect the interests of retail investors from managerial agency issues, dilution of the fund value and reducing the risks arising from the problems of free cash flow. To that extent, the Revised Listing Rules correctly identifies the requirement of ring-fencing SPAC IPO proceeds.

It is argued, however, that the provision does not provide effective protection against misappropriation and excessive diversion of funds to working capital expenses. Firstly, the method in which such funds are to be 'ring-fenced' or the level of protection offered to such funds has not been specified for the purposes of 'flexibility.' It is argued that such flexibility and vague direction to

[108] Financial Reporting Council, The UK Stewardship Code 2020, Principle 1 <https://www.frc.org.uk/investors/uk-stewardship-code> accessed on 1 August 2021.

[109] Financial Conduct Authority, *Investor Protection Measures for Special Purpose Acquisition Companies: Proposed Changes to the Listing Rules Policy Statement,* July 2021 (PS 21/10), 12.

[110] Revised Listing Rules (n 30), LR 5.6.18A (2).

[111] ibid.

ensure 'adequate protection' fails to prevent misappropriation. The Revised Listing Rules fail to mandate the use of trust structures and/or escrow accounts for the purpose of ring-fencing proceeds and leaves the same to the discretion of SPAC Sponsors. Secondly, there is no independence criteria specified for identifying third parties with whom the SPAC IPO proceeds are to be ring-fenced. The Policy Statement clarifies that such independent third parties should be 'appropriate,' experienced and a separate legal entity, free from any control or influence of the SPAC.[112] This may lead to funds being deposited and misappropriated by unscrupulous and unregulated third parties or other connected parties. Thirdly, by failing to specify a percentage threshold of IPO proceeds which can be used to fund working capital requirements, the same is left open to the discretion of SPAC Sponsors. It is argued that mere disclosure of specified amount of issue proceeds to be utilised to fund working capital needs does not provide adequate protection, as compliance of the same can be engineered through clever accounting practices.

(iv) Time-Bound Acquisition Mandate

The Revised Listing Rules provide that once admitted to listing, a SPAC must find and acquire a target within two years, subject to an extension for another year, following approval by its public shareholders.[113] A further extension of six months without shareholder approval is allowed in limited circumstances where the transaction is at an advanced stage.[114]

The introduction of a time-bound acquisition mandate is aimed to help focus managers' attention to finding an appropriate target and preventing squandering of proceeds during the 'hunt'. As argued, however, in Section III of this paper, given the current climate of many SPAC listings and few quality targets, a time bound acquisition mandate creates artificial pressure on the SPAC Sponsors, exacerbating the misaligned incentives of the Sponsors to push any deal through. This is further compounded by issues of information asymmetry and no prescription for strict due diligence of target companies, which makes monitoring by institutional shareholders or meaningful exercise of shareholders' voting rights for approving extensions more difficult. Having stated the same, the author welcomes the provision relating to further extension by six months for De-SPAC Transactions which are at an advanced stage. It is argued that such short-term flexibility prevents half-baked deals, allowing for greater diligence and scrutiny.

[112] FCA Policy Statement (n 109) 13.
[113] Revised Listing Rules (n 30), LR 5.6.18A (3).
[114] ibid, LR 5.6.18A (3)(c).

(v) Increased Disclosure and Supervision by the FCA

The Policy Statement provides for various transparency and disclosure requirements for SPACs wanting to take the alternate approach to suspension of trading, thereby increasing the flow of information to the SPAC retail investor. Firstly, a SPAC which contemplates taking the alternate approach must disclose the same in the prospectus at the point of its listing.[115] This is a mandatory requirement which must be satisfied when applying to the FCA for an exemption from the suspension of trading. Furthermore, there is a continuing obligation on the SPAC to notify the FCA if it changes or removes any of the specified investor protection mechanisms and request a suspension of trading.[116]

Secondly, the Revised Listing Rules mandate increased disclosure regarding the SPAC structure at the time of target announcement.[117]These disclosures are in addition to complying with the UK Market Abuse Regulations and the transparency rules. At the point of announcing an acquisition, to the extent possible, disclosures *inter alia* on the proposed material terms of the transaction including their effects on the shareholding of public SPAC investors, description of target's business, valuation of the target etc must be made.[118] It is argued that the introduction of such disclosure requirements will allow SPAC investors to exercise their votes more effectively.

Thirdly, in cases where a member of the SPAC's board has a conflict of interest in relation to a target or its subsidiary, the Revised Listing Rules state that the board must furnish a board statement, in sufficient time ahead of the shareholder approval vote on the proposed transaction, that the transaction is 'fair and reasonable' as far as the rights and interests of public shareholders are concerned.[119] The statement is to be backed by an 'appropriately qualified and independent adviser.' The Policy Statement leaves open the question regarding the independence and qualification of such an adviser.

Fourthly, the Revised Listing Rules provide that disclosure must be made of 'any other material detail and information that the SPAC is aware of, or ought reasonably to be aware of, about the target and the proposed deal that an investor in the SPAC needs to make a properly informed decision.'[120] It is argued that this provides for an important general information right to the SPAC retail investor, which may help reel in agency costs. The effectiveness of this right and its influence over target approvals and time extensions, will greatly depend on institutional

115 FCA Policy Statement (n 109) 27.
116 Revised Listing Rules (n 30), LR 5.6.18F.
117 ibid LR 5.6.18D.
118 ibid.
119 ibid LR 5.6.18C.
120 Revised Listing Rules (n 30), LR 5.6.18.D(2)(f).

investors. It is argued that given the minimum size threshold mandate of £100m or more, the increased stake of sophisticated institutional investors is likely to incentivise more monitoring by such investors, leading to benefits for the retail SPAC investors.

Lastly, the Policy Statement provides for a greater supervisory role of the FCA and states that SPACs which incorporate the prescribed investor protection mechanisms would have to apply to the FCA before announcing an agreed or contemplated transaction to avoid the presumption of suspension of trading.[121] At the point of listing, the Policy Statement states that the FCA shall provide comfort to the SPAC that it has met the prescribed investment protection criteria to enable it to apply for an exemption at the point of acquisition.[122] Thus, the exemption from the presumption of suspension of trading is not automatic even if a SPAC meets all the specified investor protection measures at the time of listing. The Revised Listing Rules provide that the SPAC would have to make a 'board confirmation' to the FCA that it has met all the criteria regarding investor protection at the point of listing and will continue to do so post announcement and until the completion of the De-SPAC Transaction.[123] The FCA may require the SPAC to produce evidence to support the written confirmation.

(vi) Voting and Redemption Rights

To a large extent, financial contracting ensures that modern day SPACs are more sensitive to the misaligned interests of sponsors and retail investors, than their blank-cheque predecessors. Structural and contractual safeguards prevent the risk of a bad deal being transferred to the investor. Chief among them are the contractual provisions of investor voice and exit. These coupled with the liquidity offered by the public markets ensure that retail investors have the option to walk away from a bad deal.

Accordingly, to prevent poor choice of targets and minimise conflict of interest issues, the Revised Listing Rules provide that potential acquisitions require SPAC board and shareholder approval.[124] Discussion and voting on such proposals would exclude a board member who (a) has a conflict of interest in relation to the target or its subsidiaries; or (b) is a director of the target company, its subsidiary or who has an associate that is a director of the target company or any of its subsidiaries.[125] Furthermore, the FCA introduces the all-important right for SPAC shareholders to approve the proposed acquisition through a majority voting

[121] FCA Consultation Paper (n 20) para 4.32-4.38.
[122] FCA Policy Statement (n 109) para 1.7.
[123] Revised Listing Rules (n 30), LR 5.6.18C.
[124] Revised Listing Rules (n 30), LR 5.6.18A (4) and (5).
[125] ibid LR 5.6.18A (4).

process. Under this, SPAC Sponsors are not allowed to vote on the issue of approving an acquisition, thereby avoiding issues of conflict of interest or vote manipulation (through market buy backs or private placement of SPAC shares). Derek (2007) identifies the various downsides of a blocked deal including wasted time and negotiations in hunting and structuring the potential acquisition and a blow to the reputation and ego of the SPAC Sponsor (who is left with worthless SPAC warrants).[126] It is argued that by empowering public shareholders with a right to reject potential acquisitions, issues relating to poor target choices can be reduced.

The Revised Listing Rules also state that SPAC shares should carry redemption rights, so that investors may exit their investment at any time prior to the completion of the acquisition and irrespective of whether the option holder voted in favour of the proposed acquisition.[127] This redemption right provides retail investors an important and cheap exit mechanism. When exercised, investors have a right to receive their pro rata share in the ring-fenced fund when the acquisition is completed.[128] Details of the redemption right is to be disclosed in the prospectus at the time of the SPAC IPO. The Policy Statement and the Revised Listing Rules indicate that the redemption option should specify a pre-determined strike price at which the option is to be exercised, fixed either as an amount or fixed pro rata share of issue proceeds which have been ring fenced.[129]

The power of the public investor to take control of the direction of the De-SPAC Transaction through voting and redemption rights, may lead many to believe that the modern SPACs should not be placed in the same category as its fraudulent black-cheque predecessor of the 1980s.[130] To an extent, SPAC Sponsors are maybe motivated to find quality targets driven by the fear of a negative vote on an acquisition proposal. It is argued, however, that the disciplining power of voice and exit of retail investors should not be overstated. Hirschman (1970) postulates that voice is an option for the dissatisfied only when exit is unavailable.[131] SPACs operate in an environment of high liquidity, allowing retail investors an exit if they do not like a deal. Unlike in private equity, where investors enjoy weak exit rights, investor exit rights in SPACs are strong and act, to an extent, as an alternative to the exercise of voting rights. Furthermore, for the SPAC retail investor, the ring-fencing of funds provides an attractive exit option, where

[126] Derek (n 31) 550.
[127] Revised Listing Rules (n 30), LR 5.6.18A (7).
[128] Rodrigues (n 6) 909.
[129] FCA Policy Statement (n 109) para 4.24.
[130] Derek (n 31) 550.
[131] Albert O Hirschman *Exit, Voice, and Loyalty: Response to Decline in Firms, Organisations, and States* (1st ed, Harvard University Press 1970) 34 <https://hdl-handle-net.gate3.library.lse.ac.uk/2027/heb.04043> accessed on 1 August 2021.

funds to finance the proposed acquisition are paid up front with an option to 'redeem' the same if not appealing.[132] Given the high coordination and information cost in exercising 'voice,' exit provides a cheaper route (relatively speaking) to an unhappy retail investor. Lastly, an exit reduces the funds available to finance the acquisition, making the target reluctant to go through with the deal, thereby disciplining SPAC Sponsors. It is thereby argued that in a SPAC, 'voice' of the retail investor is a residual option, whereas exit is a more powerful mechanism to deter opportunistic Sponsors.

As stated above, the issue of high information and coordination costs and a lack of effective monitoring of the SPAC Sponsor, make it difficult for the unsophisticated retail shareholder to make her voice heard. Thereby, any meaningful exercise of voting rights will be left to large and powerful institutional investors such as hedge funds. While discussing the power of voice and exit in US SPACs, it has been argued by Rodriguez and Stegemoller (2013) that the mechanisms of voice and exit created a holdout problem for SPAC Sponsors, which made it harder to get deals approved.[133] Post the financial crisis of 2007, hedge funds used this hold-out right to 'greenmail' SPAC Sponsors. Hedge funds used their voting rights to gain concessions from SPAC managers who were eager to close acquisition deals.[134] Rodriguez and Stegemoller (2013) argue that as a reaction, SPACs addressed the issue of shareholder voting right by effectively taking it away.[135] For example, in the United States, prospectuses of SPACs may not provide for shareholder approval of acquisitions and instead provide for a tender offer mechanism. Under this, the SPAC offers to buy back shares from dissenting shareholders,[136] thereby dimming the threat of a negative vote. Given that the FCA introduces a mandatory requirement ensuring shareholder approval of proposed acquisitions, such a tender offer route is unlikely to emerge in the UK. UK SPACs may still fall victim to the issue of hold-out by powerful institutional investors, who may force management to grant them concessions in exchange for votes, which may not always be beneficial for the retail investors.

V. Conclusion

The Consultation Paper and the Policy Statement recognise that the presumption of suspension of trading acted as a disproportionate barrier for both the SPAC investor and Sponsors, which kept UK from being an attractive destination for

[132] Rodrigues (n 6) 915.
[133] Rodrigues (n 6) 907.
[134] ibid 911.
[135] ibid.
[136] ibid.

SPACs.[137] In their analysis, the FCA concluded that the effectiveness of the proposed investor protection measures outweighs the advantages of the presumption of suspension of trading.

It is noted that SPAC structures have evolved overtime to imbibe various structural and contractual features that protect the retail public investor. The suggested investor protection mechanisms under the Revised Listing Rules are optional in nature and SPACs that meet the prescribed standards can apply to the FCA on a discretionary basis if the benefits of the same outweigh the costs. The changes draw from experiences in other SPAC-friendly jurisdictions during the wave of 2020 and have been drafted keeping in mind the protection of retail investors. The light-touch regulatory approach of the FCA is sensitive to the agency cost issues which have been highlighted in Section III of this paper. The Consultation Paper identified four issues which the Revised Listing Rules seek to address at various stages of the SPAC life-cycle (a) shareholder control; (b) conflicts of interest; (c) misappropriation of issue proceeds; and (d) options for issuers and increase investment opportunity for investors.[138] It is concluded that provisions such as ring-fencing of issue proceeds, increased transparency, additional disclosure requirements and introducing shareholder voting and redemption rights, goes a long way in addressing the four issues identified by the FCA and in protecting the interests of SPAC retail investors. This paper has also argued that an over-reliance on the participation and diligence of SPAC institutional investors in protecting retail investor interests is misplaced. Furthermore, it is yet to be seen whether such regulatory efforts, considering the ebbing of the SPAC wave, is nothing but an 'epic party followed by an epic hangover.'

[137] FCA Consultation Paper (n 20) 27.
[138] FCA Consultation Paper (n 20) 27.

Cambridge Law Review (2022) Vol VII, Issue 1, 57–85

Tethering the Crypto-Asset Market: The Regulation Of Stablecoins In the European Union And United States

DESIREE VAN IERSEL*

ABSTRACT

Cryptocurrencies are likely to disrupt the traditional financial system and alter how we pay for goods and services. While first-generation cryptocurrencies fail to maintain a stable value, making it a less attractive alternative to traditional money, their second-generation counterpart may fulfil the promise of digital payment.

Stablecoins can maintain stable value and therefore function as a more secure alternative. This disruptive means of payment has suddenly attracted considerable attention after the publication of the Libra (now Diem) Whitepaper. Regulators all over the world are faced with the challenge of regulating this ledger-based means of payment.

This article provides the first comparison and assessment of the EU and US proposals to regulate this technology: the Markets in Crypto-Assets Regulation (MiCA) and the US Stablecoin Tethering and Bank Licensing Enforcement (STABLE) Act. The core problems present in these proposals are highlighted and compared and possible solutions outlined.

Using the sliding scale of consumer protection and innovation as a yardstick to assess these proposals, it becomes clear that neither the EU nor the US proposals fully grasp the complexities of DLT and the reality of Stablecoins while offering a proper level of consumer protection. This article highlights the deficiencies present in these legislative instruments and proposes solutions to these core problems.

Keywords: Stablecoins; Markets in Crypto Assets Regulation; STABLE Act; Blockchain

* BA LLB LLM. I am grateful to Dr Konrad Borowicz for his insightful suggestions and feedback on earlier versions of this article, the editors of the Cambridge Law Review for their helpful comments, Wim and Monique van Iersel for their love and unwavering support and Rutger Timmer, for everything. Joke van Iersel, this one is for you. Any errors that remain are my own.

I. INTRODUCTION

Since its genesis, cryptocurrency has been lauded as a possible alternative to traditional payment methods. Cryptocurrencies are built on a blockchain and can be accessed via mobile wallets that allow users to easily send, receive and secure their tokens. Payment via blockchain has become much easier, faster and cheaper. While the most well-known cryptocurrency Bitcoin was designed to be decentralised to ensure peer-to-peer transactions without government or corporate intervention, Stablecoins move away from the libertarian roots of crypto. In contrast to cryptocurrencies such as Bitcoin, Stablecoins are pegged to an external reference point such as (a basket of) assets, fiat currency, or even algorithms. While Stablecoins may not be as true to the cypherpunk ideology of the 90s, they are better equipped to deal with the core problem that reduces the utility of first-generation cryptocurrencies: extreme volatility. Due to this difference in design, Stablecoins could, theoretically, be a viable alternative means of payment fit to launch us into the future of digital payment.

On the flip side, the rapid rise of Stablecoins can potentially impact the global financial market and global financial stability. In 2021, the market capitalisation of Stablecoins has quadrupled to more than USD 120 billion.[1] This means that right now, it is comparable to US high-yield bonds—a well-established asset class. Moreover, trading volumes have also increased exponentially, which makes exposure to spillover effects because of these (un)Stablecoins a dangerous possibility. Different from US high-yield bonds is that Stablecoins are not properly regulated. This lack of regulation exposes our global financial system to the risk of spillover effects and crises as Stablecoins have become the centre of the crypto-storm. Most of today's crypto-trading is done via Tether, the US dollar of the cryptocurrency world. If Tether were to fail, the entire crypto-market would be affected and the markets backing the token would be severely impacted.[2] The enormity of the spillover effects this would cause has already been compared to the crisis following the collapse of the Lehman Brothers bank in 2008.[3]

The risks posed by Stablecoins suddenly became clear to regulators all over the world when Facebook announced the launch of Libra (now "Diem"), a Stablecoin offered to their 2.8 billion active users. Both the US and the EU have

[1] Parma Bains and others, 'Global Financial Stability Report: Covid-19, Crypto, And Climate: Navigating Challenging Transitions', (IMF 2021).

[2] Mathijs Rotteveel and Pim Brasser, 'Op zoek naar de mysterieuze Nederlander achter de controversiële cryptomunt tether' *Financieel Dagblad* (Amsterdam, 4 January 2022); Financial Stability Board, *Assessment of Risks to Financial Stability from Crypto-Assets* (2022 FSB report) 4.

[3] ibid.

taken a stand against Diem and prevented the launch of the token in its proposed form but are now faced with the inevitable challenge of regulating the crypto-world. This boils down to the challenge of unbundling the centuries-old system of banking, money, and payments or choosing to stay behind in the innovation race. New FinTech products provide the opportunity of unbundling banks and allowing room for innovation in a centuries-old system. Laws that further entrench and bundle these three will only present bigger challenges to innovation and stop progress in payment.[4] Safely unbundling while ensuring consumer protection as well as furthering innovation is the complex balance that must be struck in the process of legislating cryptocurrencies and primarily Stablecoins.

The two frontrunners in the field recently published their legislative proposals to make up for the lack of all-encompassing crypto regulation. In December 2020, the US *Stablecoin Tethering and Bank Licensing Enforcement* ("STABLE") Act was published. The European Union published their proposed *Markets in Crypto-Assets Regulation* ("MiCA") in September 2020. These unprecedented proposals which aim to make up for the myriad of laws that are partially but never fully applicable to crypto are at the centre of this study. The key problem in any proposal aimed at regulating new technology is striking a productive balance between consumer protection and innovation. This article demonstrates that both MiCA and the STABLE Act do not strike the right balance by analysing and comparing the key components of these proposals through the balancing act between consumer protection and innovation.

II. Methodology

The most recently proposed US bill at the time of writing (the STABLE Act) and the EU Markets in Crypto-Assets Regulation were chosen for this comparative study as most Stablecoin projects originate from Europe or the USA. The proposals aim to form the core of crypto regulation and clear up the mosaic of laws that are currently (partially) applicable to crypto. For this reason, this article will not focus on the applicability of (among others) the PSD II, MiFID II, AMLD5, EMD II, and UCITSD to Stablecoins.

Instead, the focus of the discussion will be to see if the proposals provide the right balance between consumer protection and innovation. Due to the enormous risks to the global financial system, it is necessary to ensure that consumers are protected, service providers can provide a certain level of service, and issuers are regularly checked and audited. Although the importance of this component of the proposals cannot be overstated, it is also relevant to make sure

[4] Dan Awrey, 'Unbundling Banking, Money, and Payments' (2021), 110 Georgetown Law Journal 1, 5.

that it is not simply regulated out of existence due to a lack of focus on the benefits of this innovation.

The relevant dimensions of comparison are therefore consumer protection and innovation. By their very nature, these dimensions can be considered strengths when taken into account by regulators, yet the moment the balance tips mostly toward one, it becomes a weakness. Allowing room for innovation may leave the consumer vulnerable yet implementing safeguards for consumers may stifle innovation. Striking a productive balance between these dimensions is among the greatest challenges for regulators when faced with technological innovation. This has been recognised by the drafters of MiCA as well as the STABLE Act as they aim at fostering innovation while at the same time protecting consumers.[5] These dimensions and the balance that must be maintained were chosen as the core components of description and analysis in this study. The Cambridge Dictionary defines them as "the protection of buyers of goods and services against low quality or dangerous products and advertisements that deceive people"[6] and "(the use of) a new idea or method".[7]

These dimensions are subdivided into three subcategories. The dimension of consumer protection is divided into (a) the rights of token holders; (b) supervision of the token issuers and service providers; and (c) liability and enforcement of token requirements. The three subdimensions were chosen as all three aim at protecting the consumer. Innovation as a dimension is subdivided into (a) technological neutrality; (b) suitability of the proposal to regulate the technology; and (c) administrative impediments. This section is followed by a comparative analysis of the weaknesses of the proposals along these dimensions.

III. Stablecoins and Cryptocurrencies

Stablecoins are defined as "cryptocurrencies maintaining a stable value against a target price, generally US dollars."[8] These cryptocurrencies are built on a

5 Proposal for the Stablecoin Tethering and Bank Licensing Enforcement (STABLE) Act <https://tlaib.house.gov/sites/tlaib.house.gov/files/STABLEAct.pdf?utm_campaign=BitDigest&utm_medium=email&utm_source=Revue+newsletter> accessed 11 September 2021. ; Proposal for a Regulation of the European Parliament and of the Council on Markets in Crypto-Assets (MiCA), and amending Directive (EU) 2019/1937 [2020] COM/2020/593.

6 Cambridge Dictionary, 'Consumer Protection' *(Cambridge Dictionary)* <https://dictionary.cambridge.org/dictionary/english/consumer-protection> accessed 10 September 2021.

7 Cambridge Dictionary, 'Innovation' *(Cambridge Dictionary)* <https://dictionary.cambridge.org/dictionary/english/innovation> accessed 10 September 2021.

8 Marco Dell'Erba, 'Stablecoins in Cryptoeconomics. From Initial Coin Offerings (ICOs) to Central Bank Digital Currencies (CBDCs)' (2019) New York University Journal of Legislation and Public Policy <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3385840> accessed 10 September 2021, 6.

blockchain, "an open and distributed ledger ("DLT") that can (manually or automatically) record transactions between users."[9] In contrast to first-generation crypto, Stablecoins are often centralised instead of decentralised. This allows for the implementation and enforcement of a stabilisation mechanism that gives this cryptocurrency its name. Stablecoins can be subdivided into three types depending on their stabilisation mechanism, namely stablecoins that use (a) traditional collateral (off-chain backed); (b) algorithmic; or (c) crypto collateral (on-chain backed).[10]

Off-chain, fiat-backed Stablecoins are directly backed by a fiat currency or basket of currencies, as the name suggests. The issuer of this token must hold and store a reserve (or basket of currencies) to make sure that their Stablecoin is redeemable and maintains stable value. The second type is the on-chain backed Stablecoin, tokens backed by other cryptocurrencies. These fully decentralised Stablecoins do not require one central issuer to regulate the maintenance of the stabilisation mechanism which is in line with the core thought of the underlying blockchain technology.[11] On-chain-backed Stablecoins are not the focal point of this study as these tokens only move the volatility problem experienced by the first generation cryptocurrencies to the Stablecoin-level as they are backed by volatile tokens. Algorithmic Stablecoins, the third type, are also not the focal point of this study. These tokens need an Oracle to maintain the exchange rate of the cryptocurrency, more commonly referred to as exchange-rate targeting, a practice National Banks used in the past.[12] The system interferes the moment the price of the Stablecoin dips below the set amount of dollars or is worth more than the set amount. If the price dips below the set value, the amount of Stablecoins held must be decreased to maintain a stable value. The inverse is true the moment the price increases. [13] This simple mechanism has been used by central banks but was abandoned after failures in the past.[14]

This study will only focus on the regulation of (off-chain) asset- and fiat-backed Stablecoins as these are the most viable option to provide the much-needed stability as well as to allow room for consumer protection.

[9] Thibault Schrepel, 'Collusion by Blockchain and Smart Contracts' (2019) 33 (1) Harvard Journal of Law and Technology, 117, 119.

[10] Aaron Wright and Primavera De Filipi, *Blockchain and the Law: The Rule of Code*, (1st edn, HUP 2018), 10; Jeremy Clark, Didem Demirag, and Seyedehmahsa Moosavi, 'Demystifying Stablecoins' (2020) 18(1) ACM Queue, 5.

[11] ibid.

[12] ibid 14.

[13] ibid 15.

[14] ibid 14.

IV. Contrasting MiCA and the Stable Act

Both proposals have their strengths and weaknesses and can foster or stifle innovation in this field. The balancing act between safeguarding consumer protection and allowing room for innovation is among the biggest challenges for regulators when faced with (technological) innovation. If the balance tips the wrong way, it might mean taking oneself out of the race to become a market leader or allowing too much room for potential risks to global financial stability. Consumer protection and innovation together form the balance that must be kept and form an ever-present trade-off in legislation. These dimensions of comparison are used to describe and analyse the weaknesses and suitability of the proposals in the following sections. This section will first describe the definitions used in MiCA and the STABLE Act before outlining the their weaknesses relating to consumer protection and innovation.

A. Defining stablecoins

(i) Definitions in the STABLE Act

The Federal Deposit Insurance Act ("FDIA") provides the definition of a *bank* in the United States. Section 3 (1) of the Act defines banks as "(A) any national bank and State bank, and any Federal branch and insured branch, and (B) includes any former savings association."[15] These entities are engaged in "the business of receiving deposits, other than trust funds (as defined in this section)."[16] The STABLE Act proposes to amend these provisions by adding "Stablecoins issued by such bank or savings association; and" after the aforementioned clause.[17] Stablecoins are defined as

> any cryptocurrency or other privately-issued digital financial instrument that (a) is directly or indirectly distributed to investors, financial institutions or the general public; (b) is (i) denominated in United States dollars or pegged to the United States dollar; or (ii) denominated in or pegged to any other national or state currency; and (c) is issued (i) with a fixed nominal redemption value; (ii) with the intent of establishing a reasonable expectation or belief among the general public that the instrument will retain a nominal redemption value effectively fixed; or (iii) in such a manner that,

[15] The Federal Deposit Insurance Act of 1950, Pub.L. 81–797, 64 Stat. 873, s 3(1).
[16] ibid.
[17] *Tlaib* (n 5).

> regardless of intent, has the effect of creating a reasonable expectation or belief among the general public that the instrument will retain a nominal redemption value that is so stable as to render the nominal redemption value effectively fixed.[18]

(ii) Definitions in MiCA

While the STABLE Act approaches the regulation of cryptoassets narrowly, MiCA makes use of a catch-all definition of crypto-assets to make sure that those outside of the scope of one of the previously mentioned directives are covered by the framework proposed. MiCA defines crypto-assets as "a *digital representation of value or rights which may be transferred and stored electronically, using DLT or similar technology*"[19] and introduces three different regulatory sub-regimes corresponding to three sub-types of crypto-assets: *(a) Asset-Referenced tokens ("ARTs"),* 'a type of crypto-asset that purports to maintain a stable value by referring to the value of several fiat currencies that are legal tender, one or several commodities or one or several crypto-assets, or a combination of such assets'[20] *; (b) E-Money tokens ("EMTs"),* "a type of crypto-asset the main purpose of which is to be used as a means of exchange and that purports to maintain a stable value by referring to the value of a fiat currency that is legal tender"[21] *and; (c) utility tokens, a* type of crypto-asset which is "intended to provide digital access to a good or service, available on DLT, and is only accepted by the issuer of that token."[22]

Stablecoins will, dependent on their technical properties, fall either within the definition of E-Money or Asset-Referenced tokens. The third category, utility tokens, does not apply to Stablecoins but forms a catch-call clause that allows room for innovation within the field of crypto yet at the same time ensures a regulated environment. Due to the limited scope of this article, only Asset-Referenced and E-Money tokens will be discussed in-depth.

B. Consumer protection in MiCA

(i) Rights of Token Holders

The issuance of crypto-assets must be preceded by the publication of a Whitepaper, an information document detailing (among others) the project, rights and obligations concerning the crypto-asset, and the risks involved. These

[18] ibid.
[19] *MiCA* (n 5) art 3(1)(2)).
[20] ibid 3(1)(3)).
[21] ibid art 3(1)(4).
[22] ibid art 3(1)(5).

Whitepaper requirements differ to cover token-specific risks involved and grant token holders different rights dependent on the type of crypto-asset issued. Holders of E-Money tokens and Asset-Referenced tokens have many similar rights under MiCA that take into account the core difference between the two: they are issued by private parties or by credit institutions. It is surprising to note that these two regimes that together aim to regulate Stablecoins differ with regard to the right granted to holders to redeem their tokens. Issuers of ARTs have no obligation to grant holders a redemption right whereas issuers of EMTs are obligated to grant this right.[23] To ensure a certain level of protection for holders of ARTs, "mechanisms must be put in place to ensure the liquidity of the Asset-Referenced token".[24]

This means that the level of protection offered to consumer-holders is different depending on the token and its technical specification. Differentiating between EMTs and ARTs does not provide consumers with the same level of protection and may even be considered regulatory arbitrage. It could be argued that it would be beneficial to consumer protection to offer the same rights to consumers.[25] Revising the proposed framework to include an obligatory redemption right for ARTs as proposed by the ECB may, at face value, strengthen the safeguards in place for consumers. Such a revision may not have been implemented in the first place for two reasons which will be set out below.

Firstly, the core differences between these regimes are the type of party issuing the token and their respective stabilisation mechanisms. Issuers of these tokens differ as one is a private party and the other a credit or E-Money institution. The latter is somewhat similar to a bank and is subject to a number of additional prudential safeguards laid down in the E-Money Directive.[26] Such safeguards flow from the E-Money Directive as well as from MiCA. Among them are for example capital requirements, initial capital requirements, and specific insurance arrangements. These stringent safeguards ensure that issuers can cover the redemption of tokens without liquidity risk. Issuers of ARTs are private parties who, contrary to their EMT counterparts, are not subject to such extensive prudential requirements. Meeting demand for redemption of tokens may be difficult for these parties as it can massively impact the liquidity and solvability of

[23] ibid art 17(1).

[24] ibid art 35(4).

[25] Opinion of the European Central Bank of 19 February 2021 on a proposal for a regulation on Markets in Crypto-assets, and amending Directive (EU) 2019/1937 (CON/2021/4) 2021/C 152/01 (2021) 4.

[26] Directive 2009/110/EC of the European Parliament and of the Council of 16 September 2009 on the taking up, pursuit and prudential supervision of the business of electronic money institutions amending Directives 2005/60/EC and 2006/48/EC and repealing Directive 2000/46/EC (2009, OJ L 267 10.10.2009, 7, arts 3–9.

their business. Ensuring longevity and liquidity of these businesses can be facilitated by not granting a redemption right to holders.

The second and perhaps more convincing argument is best made by considering the hypothetical scenario in which an issuer of Asset-Referenced tokens is obligated to offer the right to redeem the tokens at par value at any given time. A liquidity risk will occur the moment an issuer of ARTs cannot meet the outstanding redemption requests. Such a maturity mismatch occurs when the holders cannot liquidate their tokens because the issuer does not have enough short-term assets. Traditional financial services (and credit or E-Money institutions) are subject to a number of prudential safeguards put in place by banking regulations to make sure this mismatch does not occur and redemption requests can be met. Issuers of ARTs are not subject to these safeguards, leaving them vulnerable to liquidity risk in case a large number of redemption requests are made at the same time. This risk can be mitigated and demands met through the rapid liquidation of the highly liquid financial instruments in the basket of assets stabilizing the ART (as per article 34 MiCA). Rapid liquidation of a large number of assets may result in an adverse impact on the markets of these reserve assets and spillover effects in other markets as well as negatively impact the stability of the Asset-Referenced token offered. Attempting to meet demands through rapid liquidation may destabilise the stabilisation mechanism in place and detract from the use of and trust in the money substitute.

What may therefore resemble an inconsistency in the framework governing Stablecoins may function as a prudential safeguard. This seemingly inconsistent approach should therefore not be dismissed out of hand without weighing the possible consequences first.

However, if MiCA is amended to include an obligatory redemption right for holders of Asset-Referenced tokens, as per the ECB's wishes,[27] it would be beneficial to mitigate the liquidity risk that may ensue in a manner similar to how open-end funds cover this risk. Managers of these funds can gate or suspend the redemptions until the fund can meet the requests. A similar approach should be allowed for issuers of Asset-Referenced tokens to take up a provision in their Whitepaper that functions as a gate or suspension provision, similar to those used in open-end funds. These provisions should, together with the rest of the Whitepaper, undergo scrutiny before issuance to ensure they do not impact the holders of the tokens disproportionately. Allowing the incorporation of such a provision in the Whitepaper would mitigate the liquidity risks that may occur if issuers of ARTs are obligated to grant a right of redemption. It will also cushion any spillover effects that may occur in the markets of the basket of assets stabilizing the token.

[27] *Opinion of the ECB* (n 25) 4.

(ii) Supervision

Significant Asset-Referenced Tokens ("SARTs") are supervised at the European level by the European Banking Authority ("EBA") to guarantee the same level of supervision and prevent supervisory arbitrage.[28] The supervisory regime for Significant E-Money Tokens ("SEMTs") is different, even though similar risks are cited for these two types of tokens. SEMTs are subject to a more stringent supervisory regime compared to SARTs for which no economic reason seems to exist. These significant tokens are supervised by both the EBA as well as the National Competent Authority ("NCA") and both are exclusively responsible to carry out their specific task.[29] The EBA must ensure compliance concerning custody requirements and investment of the reserve assets as stipulated in articles 33 and 34. Proper liquidity management, effective risk management, and that different Crypto-Asset Service Providers ("CASPs") can hold tokens in their custody as per article 41, the establishment and maintenance of a wind-down plan as meant in article 42, and increasing the percentage of reserve assets to be kept by the issuer as outlined in article 41(4) must be ensured. The NCA is in charge of ensuring compliance with the other obligations flowing from MiCA. Not only does this difference seem arbitrary due to the similar risks attached to these tokens, but dual supervision also has enormous drawbacks. Dual supervision by the national and the European authorities is overly complex and may lead to redundancy at the cost of overlooking other obligations. All significant tokens should be supervised at the European level to provide a level playing field as well as guarantee holders the same level of supervision.

A dual supervisory regime would also further complicate the applicable regulatory framework for issuers of significant E-Money tokens.[30] Not only would the aforementioned dual regime apply once an E-Money token is classified as significant, if the issuer is classified as a significant credit institution as per article 6(4) of the Significant Supervisory Mechanism Council Regulation ("SSM Regulation"), he would be subject to supervision by yet another authority.[31] If this were the case, the NCA, EBA, and the ECB would share the burden of supervising the issuer. This would make the supervisory regime too complex and blur the lines of competence even further. Moreover, both MiCA, as well as the SSM Regulation, require cooperation between the national and the European authorities.[32] MiCA

[28] *MiCA* (n 5) recit. 66.

[29] ibid arts 52 in conjunction with 98(4).

[30] *Opinion of the ECB* (n 25) 9.

[31] Council Regulation (EU) No 1024/2013 of 15 October 2013 conferring specific tasks on the European Central Bank concerning policies relating to the prudential supervision of credit institutions (2013) OJ L 287, 29.10.2013, 63–89, art 6(4). ; *Opinion of the ECB* (n 25) 9.

[32] ibid art 6; ibid 9.

imposes the creation of a Supervisory College consisting of among others the NCA and the EBA for a SEMT and the SSM Regulation establishes Joint Supervisory Teams established for each significant credit institution consisting of staff from the ECB as well as the NCA. To further complicate matters, the EBA's Board of Supervisors consists of the National Banking Supervisors of the Member States, that is, the NCAs. Establishing such a supervisory regime would make delineating responsibilities extremely difficult and might make the system prone to leveraging by an NCA.[33]

The dual supervisory regime as proposed in MiCA should be replaced by a similar regime as suggested for SARTs. That would create certainty and avoid regulatory arbitrage. If this approach is taken and an E-Money token issuer is classified as a significant credit institution and his tokens as significant, both the EBA and the ECB would play a role in the supervision of the issuer. Their respective responsibilities and competencies still warrant further clarification to ensure no conflicts arise and no supervisory requirements are overlooked. Ensuring the same level of regulatory supervision at the European level will offer the large group of holders of significant tokens sufficient protection and ensure compliance with the additional requirements issuers of significant tokens are subjected to.[34]

(iii) Liability and Enforcement

The regulatory regime in place for crypto-assets other than ARTs differs in many regards, but the most surprising matter relates to the approval and authorisation to issue crypto-assets. Those wishing to issue crypto-assets or EMTs (other than ARTs) do not have to wait for ex-ante approval of their Whitepaper. It merely has to be submitted and notified to the NCA 20 days before the crypto-asset or EMT is offered. Although these authorities can intervene and supervise the issuer according to article 82 of the regulation, no check beforehand takes place. This leaves the matter of accountability and enforcement regarding any misleading, incomplete, or unfair information provided in the Whitepaper to be determined *ex post* via claims for damages.

The mere *ex post* accountability and enforcement were chosen to avoid an undue administrative burden on the competent authority.[35] Yet, this approach leaves too much room for uncertainty and does not offer sufficient protection to the holders of these crypto-assets and EMTs. To create certainty and ensure a sufficient level of protection for holders, an *ex ante* system of approval and

[33] *Opinion of the ECB* (n 25) 9.
[34] *MiCA* (n 5) 68.
[35] ibid recit. 19.

authorisation is necessary—even though it might be burdensome. Especially in the crypto-sphere in which Ponzi schemes and scams still occur, such *ex ante* approval and authorisation is no unnecessary luxury to protect consumers.[36] As reviewing and approving Whitepapers can be time-consuming, it may be beneficial to allow an NCA more time to do so—a mere 20 days may not be sufficient.

Surprising to note is that MiCA provides a specific liability regime for the information presented by the issuer in its Whitepaper but does not offer an overarching European approach toward other forms of mismanagement, instability of tokens, or loss other than resulting from malfunction or hacks. Instead, articles 14 and 22 of MiCA merely state that further civil liability based on national law cannot be excluded. At face value, this may seem a proper approach that offers sufficient redress for those seeking damages. However, if one of the core benefits MiCA offers is taken into account, it does not make sense to settle this at the Member State level. As issuers of crypto-assets can get a European passport to offer their services in the Union and make use of the internal market, claiming damages at the national level would be burdensome and may not provide the same level of redress for every consumer. Regulating this at the European level may provide the consumer with compensation and a road to redress. This must be addressed in a manner different than the liability regimes currently in place at the European level such as the Product Liability Directive[37] or the E-Commerce Directive[38] as those parties issuing or offering services relating to crypto-assets do not fall within the scope of either due to the complex nature of the underlying DLT.

C. Innovation in MiCA

(i) Suitability of the Proposal to Regulate the Technology

The proposed MiCA Regulation regulates issuers and crypto-asset service providers. These parties, whether they are private parties or credit institutions as meant in the E-Money Directive, are the addressees of most provisions of the proposal. This seems suitable to regulate Stablecoins such as Tether and TrueUSD

[36] Aleksander Berentsen and Fabian Schar, 'The Case for Central Bank Electronic Money and the Non-case for Central Bank Cryptocurrencies' in Antonio Fatas (ed) *The Economics of Fintech and Digital Currencies* (CEPR Press 2019), 65.

[37] Council Directive 85/374/EEC of 25 July 1985 on the approximation of the laws, regulations and administrative provisions of the Member States concerning liability for defective products OJ L 210, 7.8.1985.

[38] Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market (2000) OJ L 178, 17.7.2000, 1–16.

as they have a centralised issuer. Yet, the DLT system has been created to cut out the middleman and to provide a decentralised means of payment. The deliberate choice to do so stems from distrust in the traditional financial system evidenced by the text in the first block on the Bitcoin blockchain stating "Chancellor on brink of second bailout for banks"- a reference to the financial crisis of 2008.[39] However unorthodox the organisational structure of these Stablecoins may be in light of the foregoing, the technological reality is that several Stabecoins are centralised while others are decentralised.

Dai for example is such a decentralised Stablecoin, issued by the MakerDAO and the Ethereum protocol. No centralised issuer can be identified as the MakerDAO and the Ethereum smart contracts cannot be considered "a legal person who offers to the public any type of crypto-assets or seeks the admission of such crypto-assets to a trading platform for crypto-assets."[40] The quoted definition aims to capture those bringing crypto-assets (in the broadest sense of the word) on the market, yet disregards the complexity and possibilities of DLT. Due to the decentralised nature of Dai, no legal entity can be considered the issuer which means, as a result, that no party is obligated to submit a Whitepaper before launch, no authorisation is needed to issue tokens, and the other supervisory and liability provisions provided by MiCA are applicable. Although recital 26 of the proposal establishes that tokens such as Dai cannot be considered as ARTs, other obligations flowing from MiCA may still apply to Dai. Yet again, this requires a centralised issuer instead of a token generated by smart contracts. Disregarding the nature of the technology underlying Stablecoins and other types of crypto-assets in this way means leaving tokens unregulated which can have a major impact on the financial system. Not taking the decentralised nature of DLT into account and the possibilities of creating a protocol generated (Stable)token makes the bespoke MiCA framework unfit to regulate the crypto-sphere as it does not encompass the technological reality of a number of these tokens.

(ii) Technological Neutrality

The principle of technological neutrality was introduced in 2002 and has been recognised as a key principle of European internet regulation in 2011.[41] This principle has been interpreted in different manners yet covers the notion that

[39] 'What is the Genesis Block?' https://coinmarketcap.com/alexandria/glossary/genesis-block accessed 10 September 2021.

[40] *MiCA* (n 5) art 3(1)(6).

[41] Organisation for Economic Cooperation and Development, *OECD Council Recommendation on Principles for Internet Policy Making* (2011) 6; Communication from the Commission to the Council, the European Parliament, the Economic and Social Committee and the Committee of the Regions. COM (99) 539 final, 10.11.1999.

regulation should not create technological silos aimed at providing a framework for one single technology. Instead of letting the technology used define the scope of regulation, a legal framework must subject the regulatory focus to the same set of principles, rules, and obligations to make sure that overarching legislation is created. This principle also ensures that a regulation such as MiCA does not become obsolete the moment DLT or crypto is no longer in use as it is focused on mitigating and steering the effects of crypto or similar technology.

The current MiCA proposal does not conform to this principle. MiCA's definition of crypto-asset refers to "using DLT or similar technology"[42] which might not be suitable to cover the crypto-asset field due to the many different crypto-projects that are built and the different designs that have already come up in DLT itself (consider for example Holochain, Polkadot, and Hashgraph). If this definition is limited to DLT or similar technologies, it might not be future-proof or provide the much-needed robust regulatory framework that is sorely lacking. Opting for a technologically neutral definition would ensure that this regulation will not be outdated in a few decades if a new type of technology is created upon which crypto-assets can rely. Adhering to this key principle to make sure that MiCA and its elaborate bespoke framework is not one of these technological silos would greatly improve the proposal as well as ensure that innovation in the field of crypto-assets can continue.

(iii) Administrative Impediments

MiCA imposes a large number of obligations on issuers and CASPs, aimed at regulating this market and protecting investors and consumers. One of the most demanding requirements is the own funds requirement in place for issuers of Significant Asset-Referenced Tokens. Tokens can be classified as significant based on their market capitalisation, customer base, interconnectedness with the financial system, the significance of cross-border activities, and the size of the reserve of assets.[43] Articles 41(4) read in conjunction with 31(1)(b) of the proposal stipulate that issuers of SARTs must have funds equal to 3% of the average amount of the reserve assets.

The most prominent Stablecoins currently in existence would immediately be classified as significant under MiCA as these already exceed 1 billion market capitalisation.[44] The requirement to maintain 3% own funds would mean

[42] *MiCA* (n 5) art 3(1)(2). ; *Opinion of the ECB* (n 25) 3.

[43] *MiCA* (n 5) art 39.

[44] Patrick Hansen, 'New Crypto Rules in the European Union – Gateway for Mass Adoption, or Excessive Regulation?' (*Stanford Law School* 12 January 2021)

maintaining a massive amount of own funds which will only increase when the token becomes more successful and mainstream. This will effectively harm the innovation and progress of Stablecoins as alternative means of payment.

D. Consumer protection in the STABLE Act

(i) The Rights of Token Holders

Under the Act, token holders are guaranteed a redemption right at par value. This guarantees security for holders as they can always get the amount of US Dollars they invested in the Stablecoin back from the issuing party. This issuer must be an insured depository institution and member of the Federal Reserve which also provides the holder of the tokens with an additional safeguard in the form of deposit insurance. Traditionally, this would only apply to the holder of a bank account for a maximum amount of USD 250,000. Deposit insurance guarantees account holders this amount in case of bank failure.[45] This fund and the fact that issuers must be insured depository institutions provide holders with the certainty that even in case of bank failure, they will be able to redeem their tokens for dollars. Holders are also protected through rules and standards set by Federal Banking Agencies as they are charged with setting the appropriate standard for capital adequacy, leverage and liquidity. As long as these agencies create a level playing field regarding the prudential safeguards and allow some room for Stablecoin-related activities, consumer-holders will benefit.

As the STABLE Act aims to embed the issuance of Stablecoins in the existing frame of bank legislation, the further lack of additional rights for token holders flowing from the Act seems sound. Holders are already protected by the myriad of obligations and rights flowing from the patchwork of banking legislation including privacy laws and bank secrecy and are essentially not treated any different than holders of bank accounts.

(ii) Supervision

Section 1(f) concerning oversight by Federal Banking Agencies stipulates that all insured depository institutions engaged in Stablecoin-related activities are supervised by a Federal Banking Agency. The Act does not specify which Agency is tasked with this supervisory role which must be done if this proposal ever becomes law. No Agency is created to take up this role which also means that this

<https://law.stanford.edu/2021/01/12/new-crypto-rules-in-the-eu-gateway-for-mass-adoption-or-excessive-regulation/> accessed 11 September 2021.

[45] Federal Deposit Insurance Act, 12 U.S.C. 1813, s 11(E).

forms an additional competence next to the general supervisory obligations relating to the business of the institution not related to Stablecoins.

Due to the complex nature of DLT, Stablecoin, and related instruments, it may be beneficial to create an Agency tasked with supervision of Stablecoin-related matters and institutions to ensure that this complex instrument and the underlying technology are subject to proper oversight.

(iii) Liability and Enforcement

As the Act essentially aims to implement the issuance of Stablecoins and activities related to it in the existing legal framework applicable to banks, it was expected that the Act itself would not contain a separate liability and enforcement regime. If claims for damages or taking an insured depository institution to court are not contractually excluded, it would be possible to do so.

E. INNOVATION IN THE STABLE ACT

(i) Technological Neutrality

The Act does not provide a complete framework aimed at regulating the crypto-asset market in the USA as MiCA does for the EU. Instead, it aims at regulating Stablecoins and defines them as "cryptocurrency or other privately-issued digital financial instrument (…)"[46]. This definition takes the technological reality into account that, DLT-based cryptocurrencies may not exist in the future if a more suitable technology is created. The longevity of the Act was ensured by adding the last clause. Although the proposed Act is neutral as it does not refer to the ledger technology underlying cryptocurrencies, it is not neutral in itself as it only refers to tokenised instruments denominated in or pegged to a currency with a nominal redemption value.[47] In American parlance, "regulation should not prejudice technological choices, by picking the winners and the losers".[48]

As expected, Stablecoins are the biggest losers. The strict regulation does not provide a framework for FinTech products as it only entrenches the existing banking system. Consider the landscape for other non-bank FinTech companies such as Venmo, PayPal, Cash App and Stablecoins. These platforms offer alternative non-bank payment platforms and are illustrative of development

[46] *Tlaib* (n 5) s 3(aa)(1), 8.

[47] ibid s 3(aa)(1)(B).

[48] Winston J. Maxwell and Daniel L. Brenner, 'Confronting the FCC Net Neutrality Order with European Regulatory Principles' (2012), 1 Journal of Regulation and Compliance, 6.

toward unbundling banking, payments and money.[49] Unsurprisingly, the US proposal does not allow such unbundling as it places issuers in the pre-existing banking framework. What is more, this Act only regulates Stablecoins and does not include other providers of payment platforms. It allows room for these platforms to thrive in a relatively unrestricted manner. The reason for this may be that providers of these platforms still rely on traditional banks to send and receive payments. These payments will later end up in the traditional bank account of the customer of the platform service. This shows that banks still have the power. Although Stablecoins are, in their very nature, similar instruments, this lack of a connection to a traditional bank may be the core concern. Stablecoins, as well as Cash App and Venmo provide dollar-denominated liabilities. These liabilities only differ in their embodiment: a token.[50] These tokens represent a claim on dollars held in a bank by the issuer of the Stablecoin and are not inherently different from mobile banking applications such as Cash App and Venmo. Both applications are peer-to-peer mobile payment apps that allow users to link their account to their bank account.[51] The user can send and receive money through this application which acts as a middleman between the banks of the sending and receiving parties. Stablecoins do not differ from this core set-up as the DLT-based token forms a claim on the dollars held by the issuer in a bank account, which, on the side of the issuer, is a liability. The mere manifestation of the liability relies on the ledger-based token while the liability in itself remains the same.[52] Perhaps the key difference between these platform payment service providers is the proximity to a traditional bank account. Reliance on the bundled system of banking, payments and money will enhance the role of banks while at the same time picking the winners and losers in FinTech—simply based on their proximity to the traditional system.

The similarities outlined above raise the question as to why the Act merely focuses on the regulation of Stablecoins instead of attempting to cover FinTech dollar-denominated liabilities as a whole. The proposal would bring Stablecoins and their issuers under the purview of the FDIC and scope of the FDIA which would significantly hamper innovation in this sphere. Moreover, it arbitrarily regulates dollar-denominated liabilities in token form. Picking the technological

[49] *Awrey* (n 4) 5.

[50] Larry Cermak, Lars Hoffmann, and Mike Rogers, *Stablecoins: Bridging The Network Gap Between Traditional Money and Digital Value* (*The Block Research* 2021) 59-61; Peter van Valkenburgh, 'The Unintended(?) Consequences of the STABLE Act' (*Coindesk* 3 December 2020) <https://www.coincenter.org/the-unintended-consequences-of-the-stable-act/> accessed 11 September 2021.

[51] *Awrey* (n 4) 5.

[52] ibid.

winners by creating a regulatory silo such as this Act limits innovation, is not neutral in its focus, and displays regulatory arbitrariness.

(ii) Suitability of the Proposal to Regulate the Technology

The key weakness of the proposal with regard to innovation is a misunderstanding of how DLT-based instruments work. The key characteristic, decentralisation, makes the regulation of this technology very complex as one cannot simply regulate one centralised legal entity and thereby regulate the technology. The distributed nature and code underlying the ledger allow the issuance of tokens through the protocol or smart contracts instead of by one centralised entity.

The proposal in its current form assumes the existence of a centralised issuer that would be suitable to regulate several Stablecoins currently in existence, such as Tether. Tether Limited is the centralised party responsible for the issuance of the token and maintenance of the reserve of assets. This party could, theoretically, apply for a banking license and conform to the obligations as laid out in the Act. Yet the moment the technological setup of a Stablecoin is not as clear-cut and straightforward as Tether's, the Act falls flat. The distributed nature of the Ledger technology used for cryptocurrencies was not taken into account in this proposal nor was the possibility of Stablecoins generated by a Decentralised Autonomous Organisation ("DAO"), a decentralised community that makes use of a blockchain to register its (financial) interactions and is able to generate its own tokens. MakerDAO's multi-collateralized Dai Stablecoin system is an example of such a token that aims to maintain a stable peg yet is not governed or managed by one single party. This Stablecoin with a 24-hour trading volume of USD 460.309.862 is one of the products of MakerDAO, an open-source software placed on the Ethereum blockchain as a Decentralised application ("Dapp") and is governed and created by the "Maker" DAO in a decentralised manner by all holders of the Dai tokens.[53] This MakerDAO allows anyone to generate tokens named "Dai" "by leveraging Ethereum as collateral through smart contracts known as Collateralized Debt Positions" and maintains a soft peg to the US dollar.[54] Stablecoins such as Dai would not be suitably regulated by the Act as no single centralised issuer can be identified. Regulating the issuer and enforcing *ex ante* approval before issuance in the manner proposed is impossible for those Stablecoins issued through a DAO or code (for example, a protocol or smart contracts). The complexity and core characteristic of the underlying technology

[53] CoinMarketCap, 'DAI' *(CoinMarketCap)* <https://coinmarketcap.com/currencies/multi-collateral-dai/> accessed 12 August 2021.

[54] MakerDAO, *The Maker Protocol: MakerDAO's Multi-Collateral Dai (MCD) System*, (Whitepaper) 2.

are disregarded in the Act which makes it unsuitable to effectively cover and regulate DLT-based instruments.

Even though the MakerDAO and the Ethereum smart contracts cannot be regulated in the manner proposed in the Act, its provision on Stablecoin-related commercial activities can still impact the DAO and the Ethereum blockchain. The Act prohibits "any person to issue a Stablecoin or Stablecoin-related product, to provide any Stablecoin-related service, or otherwise engage in any Stablecoin-related commercial activity, including activity involving Stablecoins issued by other persons"—without prior authorisation.[55] This broad provision would, as a result, prohibit all participation in Stablecoin-related activities, including the MakerDAO and the Ethereum blockchain by parties without prior authorisation to engage in these activities.

Not only would the MakerDAO be prohibited in itself as its participants and its smart contracts issue Stablecoins, running a node on the Ethereum blockchain that run the smart contracts and ensure block creation will be prohibited too. The nodes on the Ethereum blockchain occupy themselves with block creation and verification. These blocks consist of several transactions and smart contracts govern the condition(s) whereby a transaction can take place. In case a block would contain a smart contract or transaction related to Dai or MakerDAO, the node occupied with verification and creation of the block could be penalised as he does not have prior authorisation to be involved in activities involving Stablecoins issued by others (in this case, the MakerDAO).

Effectively, the STABLE Act would make it illegal for a node to participate in the Ethereum blockchain. It is impossible to select the transactions and smart contracts a node wishes to validate which, in light of this proposed Act, would be an enormous risk if one is not authorised. A node cannot pick and choose but validates all transactions or none at all. If the Act were to become law, it would mean that nodes on the Ethereum blockchain would have to cease validating transactions out of fear that one of them may relate to Dai or Maker.[56] These severe consequences of this broad prohibition will therefore not merely target Stablecoins but stifle innovation in the DLT sector. This Act may aim to promote innovation yet seems to miss its mark through a misunderstanding of DLT and the issuance of Stablecoins through a DAO, smart contract or protocol.

(iii) Administrative Impediments

As described above, the Act would require an issuer of Stablecoins to obtain a banking charter, become a member of the Federal Reserve, and an insured

[55] *STABLE Act* (n 5) s 52(2)(a)(2), 12.
[56] *van Valkenburgh* (n 50).

depository institution placed under the purview of the Federal Deposit Insurance Corporation ("FDIC"). These are not mere administrative burdens to bear for issuers wishing to comply with the proposed legislation, but massive financial burdens to bear. If an issuer would, for instance, attempt to obtain a banking charter in the state of New Jersey, they would have to pay a non-refundable filing fee of $15,000 to merely have his application considered.[57] The issuers of the largest Stablecoins currently in existence will probably be able to afford this fee but the smaller players will most likely not be able to meet this financial threshold. The Act does not provide smaller actors with any form of regulatory sandbox from which they could benefit. Instead, this requirement is placed at the forefront of the Act which will not allow small (future) disruptive players a chance to innovate and grow. This lack of regulatory flexibility and the high financial burden does not benefit this exponentially growing market which, in May 2021, broke 100 billion US dollars.[58] Administrative impediments as created by this proposal will, if adopted, likely result in a decrease of United States-based crypto innovation, result in innovation migration, denomination in other currencies, and deflate US competitiveness in the innovation race. If adopted in its current form, the United States would be acting to its detriment.

V. Innovation and Consumer Protection: MiCA and the Stable Act

Striking a productive balance between consumer protection and innovation is essential to a legislative proposal aimed at regulating a FinTech product. Both MiCA and the STABLE Act aim to balance both dimensions to ensure that fostering innovation does not open the gates to abusive practices while at the same time making sure that safeguarding consumers does not stifle innovation.[59] The analysis and description provided above show that both proposals have not successfully struck this balance. They lean too much toward the consumer protection dimension which could massively impact innovation in the crypto-sphere as well as the competitiveness of the United States and the European Union on the FinTech/DeFi market. The imbalance will be illustrated further by means of the two dimensions and the weaknesses present in the proposals. For the sake of clarity, E-Money Tokens and Asset-Referenced Tokens under MiCA as well as

[57] State of New Jersey Department of Banking and Insurance, *Requirements for Organizing a New Jersey State Chartered Bank or Savings Bank*, NJAC 3:1-2.2a(4).

[58] Joe Weisenthal, 'The US$100 billion stablecoin question' (*BnnBloomberg* 28 May 2021) <https://www.bnnbloomberg.ca/the-100-billion-stablecoin-question-1.1609865> accessed 11 September 2021.

[59] *MiCA* (n 5) para 1; *Tlaib* (n 5).

Stablecoins under the STABLE Act will collectively be referred to as Stablecoins, cryptocurrencies maintaining a stable price against a currency. If the separate provisions on EMTs, ARTs or Stablecoins are discussed, the distinction will be made clear.

A. Consumer protection

Under MiCA, a CASP custodian is "liable to their clients for loss of crypto-assets as a resulting [sic] from a malfunction or hacks up to the market value of the crypto-assets lost."[60] The CASP is liable for any damages, including ICT-related incidents such as cyberattacks, malfunctions, and theft.[61] This strict approach, although beneficial to consumer-holders, shows that CASPs are regulated in a strict manner unsuitable to foster innovation and provision of these services. Incidents as such are commonly considered force majeure events that cannot be considered a ground for a damages claim. Moreover, it is not in line with the approach generally taken for depositaries and custodians of transferrable securities. Article 19 of Regulation 2016/438 supplementing the Undertakings for Collective Investment in Transferable Securities Directive ("UCITS Directive"), states that a depositary/custodian is not liable for damages in what are traditionally considered to be force majeure events.[62] As long as the custodian can prove that the damages occurred due to an external event that could not have been avoided even though reasonable efforts were made to prevent such an event from resulting in loss, they are not held liable. Surprisingly, this standard is not implemented for their crypto-asset counterparts who essentially provide the same or similar services. If a crypto-asset custodian makes reasonable efforts to protect himself against ICT-related incidents and it happens nonetheless, they should not be held liable for all damages resulting from it. All damages arising from incidents beyond their reasonable control must be limited and this matter must be approached in a manner similar to the commonly accepted approach for force majeure events.[63] This strict approach toward liability for what could be qualified as force majeure events may make crypto-custodianship unattractive and even incredibly risky

[60] ibid art 67(8).

[61] ibid recit. 59.

[62] Commission Delegated Regulation (EU) 2016/438 of 17 December 2015 supplementing Directive 2009/65/EC of the European Parliament and of the Council with regard to obligations of depositaries (2009) OJ L 78, 24.3.2016,11–30, art 19; Dirk Zetzsche and others, 'The Markets in Crypto-Assets Regulation (MICA) and the EU Digital Finance Strategy' (2020) 2020/77 European Banking Institute Working Paper Series <https://ssrn.com/abstract=3725395> accessed 10 September 2021, 20.

[63] ibid.

business.[64] Instead, MiCA should conform to the UCITS Directive's approach and limit liability for force majeure events. Mitigating this risk for CASPs would level the playing field between them and their traditional colleagues as well as stimulate innovation in the provision of crypto services.

Contrary to the MiCA regime, the STABLE Act does not include a separate liability provision. As the Act aims to implement these tokens in the banking system, it is practical not to create an additional obligation specific to these tokens and their issuers. What must be noted is that many insured depository institutions now include a mandatory arbitration clause in their contracts, preventing consumers from going to court to claim damages as well as preventing class-action suits.[65] Claiming damages from insured depository institutions issuing Stablecoins will therefore be a difficult if not impossible reality. The debate on the appropriateness of these clauses is an entirely different topic altogether but must not be forgotten in the evaluation of the rights of consumers under the STABLE Act.

While MiCA allows private parties to issue ARTs, the STABLE Act obligates an issuer to become an insured depository institution—a type of institution that already covers the risk of bank failure and illiquidity through their backing by the faith and credit of the US government.[66] These institutions are required to redeem the Stablecoins at par value at any point in time. MiCA, as mentioned earlier, only grants this right to holders of EMTs. MiCA does not provide a mandatory right to redeem ARTs and while it may be argued that this does not offer sufficient protection to consumer-holders, this provision prevents liquidity squeeze and spillover effects in other markets. Tokens such as Stablecoins are vulnerable to bank runs and the only way to prevent liquidity problems from occurring is by establishing a gating provision or creating a lender of last resort able to meet all redemption demands as was done in the STABLE Act.[67] Although MiCA allows room for private parties to issue Stablecoins and does not contain a mandatory right of redemption for ARTs, if this right were to be implemented in the proposal as per the ECB's wishes, it must be supervised, guided, and gated properly to prevent liquidity risks. If not done properly, history will repeat itself and test if the tokens are properly backed once a bank run or confidence crisis comes up. Lessons

64 ibid.

65 *Discover Bank v Superior Court* [2005] 113 P.3d 1100, 36 Cal. 4th 148, 30 Cal. Rptr. 3d 76. ; J. Maria Glover, 'Beyond Unconscionability: Class Action Beyond Unconscionability: Class Action Waivers and Mandatory Arbitration Agreements' (2006) 59(5) Vanderbilt Law Review <https://scholarship.law.georgetown.edu/cgi/viewcontent.cgi?article=2888&context=facpub> accessed 11 September 2021.

66 Andrés Velasco, 'Preventing a Stablecoin Liquidity Crisis' (*Project Syndicate* 2 August 2021) <https://www.project-syndicate.org/commentary/preventing-a-stablecoin-liquidity-crisis-by-andres-velasco-2021-08> accessed 11 September 2021.

67 ibid.

must be learned from the case of the Argentine peso in the 90s which was supposedly backed by the US dollar. The peso was not fully backed which resulted in the collapse of the Currency Board charged with the maintenance of the peg and even the Argentine government.[68] If a Stablecoin (in the form of an (S)ART) becomes too big to fail without proper gating or a lender of last resort structure, consumers will not be sufficiently protected and the consequences for the (global) economy will be incalculable. The MiCA proposal in its current form adequately tackles this issue and changes in this structure to allow holders of ARTs the same rights as holders of EMTs should not be made without first considering the major risks and mitigation thereof.

Both MiCA and the STABLE Act leave a lot to be desired when it comes to supervision of Stablecoins. While the European proposal consists of an elaborate supervision regime including Supervisory Colleges, the STABLE Act hardly regulates supervision at all. Neither strikes the right balance between protecting consumers through suitable supervision while creating clarity for these issuer-institutions as to what rules they have to adhere to and under whose supervisory authority they are placed. While MiCA must reduce the complexity of their supervisory system, the Act, on the other hand, would benefit from the further specification of the Federal Banking Authority charged with oversight. While some may say that the Office of the Comptroller of the Currency ("OCC") should be charged with this task, it has become clear that Tlaib, García, and Lynch did not have them in mind for this task as they wrote a pointed letter in response to the OCC's recent plans to offer special-purpose payment charters. It was clear that they were concerned about the OCC's overreach their letter stated that

> The decisions of your agency have the potential to adversely affect banking and financial activities well beyond your jurisdiction. In particular, decisions regarding the classification and regulation of "crypto assets" and crypto-related payments services may have secondary effects on the entire hierarchy of financial assets denominated in U.S. dollars, as well as the more traditional means by which retail and wholesale payments are made in the United States and abroad.[69]

[68] Letter to the OCC on Fintech Charters (10 November 2020) <https://tlaib.house.gov/sites/tlaib.house.gov/files/Letter%20to%20the%20OCC%20on%20Fintech%20Charters_Tlaib_Lynch.pdf> accessed 11 September 2021, 2.

[69] Michael Bleaney, 'Argentina's Currency Board Collapse: Weak Policy or Bad Luck?' (2004) 27(5) The World Economy <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=538213> accessed 11 September 2021.

The STABLE Act would benefit from some elaboration or clarity as to what agency will be involved. Among eligible agencies, the one most likely to be involved based on the Gramm-Leach-Bliley Act would be the FTC. Its involvement would ensure that financial institutions can explain how they share and protect customer data. Another agency that is likely to be involved is the Consumer Financial Protection Bureau ("CFPB"), which came into existence through the Dodd-Frank Wall Street reform. This agency aims to protect consumers against deceptive practices by financial institutions through supervision and enforcement.[70] This patchwork of supervisory financial agencies, although typical for the United States' financial oversight regime, may not be beneficial to supervise the issuers of Stablecoins due to the complex nature of the underlying technology. Creating a clear supervisory overview or reducing the number of agencies involved would benefit these agencies, the issuers themselves, and consumers as they can rely on a cooperative collaboration between agencies. Mitigating these weaknesses in both proposals would strengthen the safeguards they aim to offer for consumers while maintaining the balance needed to allow room for innovation.

B. Innovation

The core weaknesses of both proposals are also present in the innovation dimension. Regulating DLT in a way that encompasses all the core characteristics and complexities of this technology seems to have been the greatest challenge for the lawmakers. Both proposals take a stringent approach which still leaves a lot to be desired with regard to the regulation of DLT.

In MiCA, the complexities of DLT are simply dealt with by stating that algorithmic Stablecoins that aim to maintain a stable value through a protocol cannot be considered ARTs.[71] The proposal does not include the possibility of decentralised, protocol-generated Stablecoins pegged to a single currency. While it does not regulate these tokens as Stablecoins, obligations for crypto-assets other than ARTs or EMTs may still apply. While it may be beneficial for consumers that even these tokens do not escape the reach of MiCA, it is problematic that holders are not granted the same level of protection as EMT or ART holders. This discrepancy seems to stem from an inadequate understanding of protocol-generated tokens backed by currency or a basket of assets. Lack of centralisation should not mean a lack of suitable legislation to cover technological progress. Regulating this type of token will help promote innovation while protecting consumer-holders. While MiCA does not offer sufficient consumer protection for

[70] Dodd-Frank Wall Street Reform and Consumer Protection Act, Pub. L. No. 111-203, § 929-Z, 124 Stat. 1376, 1871 (2010), Title X.; Gramm-Leach Bliley Act, Pub.L. 106–102 (1999), Title II s 201.

[71] *MiCA* (n 5) art 26.

holders of these tokens, the STABLE Act does not fare any better with regard to innovation. Mitigating these risks and addressing this imbalance can only be done by engaging with the technology, its complexities, and the experts in the field and amending the legislation accordingly. Creating room for a regulatory sandbox to gain insight into the way this technology works may be a good starting point for both the US as well as the Union. The EU has taken steps to include a pilot regime for DLT-based instruments. This regime allows DLT market infrastructures to be temporarily exempt from specific requirements of the EU financial legislation to provide the European Securities and Markets Authority ("ESMA") with a chance to gain experience and insight into the risks, technology, and opportunities of DLT.[72] The EU is already moving in the right direction and will hopefully implement the lessons learned from this pilot regime in the MiCA framework. The US would benefit from a similar approach and should take note.

The point raised regarding technological neutrality also warrants further analysis. The core problem visible in MiCA lies with the central definition of crypto-assets, which was not formulated neutrally. The STABLE Act, on the other hand, has bigger problems to tackle. While the definition used for Stablecoins does not include an explicit reference to DLT, it mentions cryptocurrencies. These are inherently ledger-based which makes the definition not entirely neutral. However, it also refers to other privately-issued digital financial instruments, providing room for non-ledger based instruments. The definition used incorporates other instruments as such and does not make reference to similar technologies or ledger-based instruments, thus ensuring its adequacy to regulate future innovation in the DeFi sphere as well. What is problematic about the STABLE Act is not the definition in itself but the fact that it does not apply to cryptocurrencies that do not fit the definition of Stablecoin. The US does not have one overarching DLT framework in place, even though Congress has been urged to create one on numerous occasions and an increasing number of states have adopted DLT-related resolutions and acts.[73] Merely regulating Stablecoins is no longer sufficient and leaving out other dollar-denominated liabilities clearly shows a preference toward instruments that are not DLT-based. Changing this to an overarching framework would be beneficial for innovation as well as for consumer protection. Surprisingly, it seems as though this thought has finally been put to action as in

[72] Proposal for a Regulation of the European Parliament and of the Council on a pilot regime for market infrastructures based on distributed ledger technology [2020] COM/2020/594 final, recit. 5.

[73] Kevin Helms, 'SEC Urges Congress to Pass Cryptocurrency Legislation to Protect Investors' (*Bitcoin* 28 May 2021) <https://news.bitcoin.com/sec-congress-pass-cryptocurrency-legislation-protect-investors/> accessed 11 September 2021; Christopher Giancarlo, 'Written Testimony of J Christopher Ciancarlo, chairman of the CFTC before the Senate Banking Committee' <https://www.banking.senate.gov/imo/media/doc/Giancarlo%20Testimony%202-6-18b.pdf> accessed 11 September 2021.

May 2021, Congressman Beyer introduced the *Digital Asset Market Structure and Investor Protection Act*, an overarching regime for digital assets.[74] This Act allows the Federal Reserve to issue their own Central Bank Digital Currency ("CBDC") and requires the SEC and CFTC to delineate the crypto-assets that fall within their respective purview. It remains uncertain if it will gain enough traction yet seems a step in the right direction.

As stated before, both MiCA and the STABLE Act create several requirements for their issuers and service providers. Under MiCA, issuers of significant tokens will have to maintain 3% of the average value of the reserve assets in own funds. To illustrate the impact of this requirement, the 3% obligation will be applied to Tether. Tether would, after the enactment of MiCA, be classified as significant. As mentioned above, this would also mean that the issuer of Tether, Tether Limited, must abide by the 3% rule of articles 41(4) and 31(2)(b). This rule requires Tether to maintain USD 1,509 billion in own funds as in April 2021, USD 50.3 billion backed Tether.[75] This is disproportionate toward those issuing significant tokens and cannot be used to regulate the crypto-sphere. If adopted, this requirement will essentially make it impossible to meet the demands imposed on issuers of significant tokens. Such a high threshold would effectively kill significant Stablecoins as issuers can hardly meet these financial requirements. The STABLE Act does not fare much better as it requires issuers of Stablecoins to become banks. This includes obtaining a banking charter, having sufficient capital to support its risk profile, and starting capital. Filling a request for a banking charter alone will already cost more than $15,000 and starting capital needed can be over $20 million.[76] Moreover, the bank must meet the minimum capital requirements in place dependent on its tier. Fitting Stablecoins into the existing framework would create too big an administrative burden as well as pose a massive financial burden. The most relevant Stablecoins currently in existence may be able to meet these financial requirements yet they may get entangled in the STABLE Act's strict approach towards all those engaged in Stablecoin-related activities.

Zooming out, it is evident that both MiCA and the STABLE Act exacerbate the bundled banking system without questioning the underlying system of money, payments, and banking. The STABLE Act proposes to bring the threat of this new, unbundled means of payment within the purview of banks while MiCA allows some unbundling to take place with regard to the provisions on issuers of (S)ARTs.

[74] Proposal for a Digital Asset Market Structure and Investor Protection Act <https://beyer.house.gov/uploadedfiles/beyer_028_xml.pdf> accessed 11 September 2021.

[75] Olga Kharif, 'Crypto's Shadow Currency Surges Past Deposits of Most U.S. Banks' (*Bloomberg* 1 May 2021) <https://www.bloomberg.com/news/articles/2021-05-01/crypto-s-shadow-currency-surges-past-deposits-of-most-u-s-banks> accessed 11 September 2021.

[76] *State of New Jersey Department of Banking and Insurance* (n 57).

Their E-Money counterparts, (S)EMT issuers, are still regulated in a similar manner as banks are. (S)EMT issuers are obligated to become credit institutions under the E-Money Directive and are placed within a similar regime as traditional banks. This indicates that both proposals privilege and protect the system currently in place and allow it to maintain an advantage over FinTech. The law serves as a means to exacerbate the existing business and setup of banks, however inefficient and not inclusive they may be. To illustrate, banks are the sole claimants on the Federal Reserve or the ECB, thereby providing a safety net in case of crises or bankruns that Stablecoin-issuers would not have.[77] In case a run on Stablecoins happens, the issuer will almost always face bankruptcy. By design, banks have an enormous advantage compared to other non-bank entities. Secondly, banks are the only entities in the US permitted to become members of the Federal Reserve and in the EU of the ECB. This legal arrangement only further exacerbates the entrenchment of traditional banks and makes them almost impossible to compete with. Opening up membership or a specific form of membership to non-bank entities would help issuers of Stablecoins (or other FinTech products) by providing a 'lender of last resort' structure. It would ensure the protection of consumers in case of crises and ensure that new players on the market who only wish to provide either money, payments or banking have room to innovate. This would decrease our reliance on banks and provide access to faster, more inclusive and less costly means.[78]

The last observation regarding the STABLE Act relates to innovation as well as consumer protection. One of the core aims of the STABLE Act is to protect vulnerable communities and low- and middle-income households from bad actors as well as tackle the barriers presented by traditional financial services and institutions. Although it is acknowledged that the underlying technology is different and presents unique challenges, the drafters claim that the core fair lending risks are the same as in the past.[79] The drafters addressed these risks in the Act by placing Stablecoins in the pre-existing financial legal framework. Instead of breaking down the barriers in place that make it difficult for members of these communities to gain access to financial institutions, the drafters placed this alternative instrument in the system that created these barriers in the first place. As elaborated upon in the 2019 FDIC survey of unbanked and underbanked households, 7.1 million households were unbanked in 2019.[80] One of the most

[77] *Awrey* (n 4) 24.

[78] ibid 37.

[79] Rashida Tlaib, 'The Stablecoin Tethering and Bank Licensing Enforcement Act: One pager' (*Tlaib House* 2020) <https://tlaib.house.gov/sites/tlaib.house.gov/files/STABLE_Act_One_Pager.pdf> accessed 10 September 2021.

[80] Federal Deposit Insurance Corporation, *How America Banks: Household Use of Banking and Financial Services* (2019 FDIC Survey), 1.

cited reasons for being unbanked is the costs involved in opening and maintaining a bank account. Bank account fees, minimum deposits and minimum balances all amount to millions of households without access to financial services, loans, lines of credit, and savings accounts. The (perceived) lack of access to these institutions and services as well as distrust in the system lies at the core of this problem.[81] Instead of truly solving the core problem underlying these barriers to access, the STABLE Act merely proposes to make Stablecoins part of the entrenched bundled banking structure that results in vulnerable communities looking elsewhere to have their financial servicing needs met.[82] This Act will not help vulnerable (unbanked) communities at all as it will merely result in maintaining the status quo with regard to consumer protection and it will further hamper innovation.

VI. Conclusion

It has come to light that both EU and US regulators aim to regulate Stablecoins in an entirely different manner even though they share the same concerns regarding financial stability, consumer protection, and innovation. While the STABLE Act solely focuses on Stablecoins and places these tokens in the existing financial system, MiCA aims to provide a bespoke framework for all crypto-assets. Both approaches, however different they may be, do not strike a productive balance between innovation and consumer protection and do not take the complexities of this technology into account.

The balance has tipped too far to the side of consumer protection and the importance of fostering innovation in this field seems sidelined by regulators. Both proposals would benefit from rebalancing the scales to foster innovation while at the same time offering sufficient protection to consumers. Involving subject-matter experts as well as creating regulatory sandboxes should be considered in order to help further the understanding and proper regulation of this technology. It is necessary to take some weight off the consumer protection side by aligning the obligations for issuers and service providers with common practice and (liability) standards in the field of traditional financial services. Furthermore, this dimension would benefit from the creation of a clear and comprehensible framework of supervision that does not blur responsibilities between agencies and is not prone to leveraging by a single actor. Mending these core issues and rebalancing these proposals can help prevent outward innovation migration while at the same time ensuring consumer protection. Moreover, striking the right balance will help ensure that this innovative instrument can be used safely without further concerns for global financial stability. As the Stablecoin market has

[81] *Tlaib* (n 5).
[82] ibid.

quadrupled in 2021, it is necessary to take action now and ensure that these new financial instruments are properly regulated. Rebalancing these proposals and allowing room to unbundle the system of money, banking, and payments could also kickstart the secure trade and use of Stablecoins, ensure US and EU competitiveness on this market, foster innovation in the financial sector and allow it to step out of the regulatory shadows. Perhaps more importantly, it could help ensure that financial stability will not be at risk in the scenario that one of these Stablecoins becomes the centre of the crypto-trading system and launches us into the future of payment.

Cambridge Law Review (2022) Vol VII, Issue 1, 86–120

Finding a Home: The Development of Himalaya Clauses in England and Canada

MARY PPASIOU*

ABSTRACT

The protection of third parties under a contract of carriage of goods by sea is the result of an incremental change in the law of England and Canada. Himalaya clauses in contracts of carriage have been devised to extend the protection of the carrier under the contract to its servants, agents, and independent contractors. However, the entitlement of third parties to rely on Himalaya clauses has been highly contentious in English and Canadian law because of the tension between such clauses and the traditional privity rule. Prior to the enactment of the Contracts (Rights of Third Parties) Act 1999 in England and the establishment of the principled exception to privity in *London Drugs Ltd v Kuehne & Nagel International Ltd* [1992] 3 SCR 299 in Canada, courts had allowed the enforcement of Himalaya clauses on legal bases such as agency, unilateral contract, or broad exemption clauses in order to circumvent privity. Even though these legal bases have been criticized for their artificiality, English and Canadian judges have been willing to allow the contractual protection of third parties for purposes of commercial practicality. This article argues that Himalaya clauses have not yet found their final place in English and Canadian law. In doing so, it critically examines the historic development of Himalaya clauses in England and Canada and discusses some theoretical and practical concerns arising from their enforcement in the shipping context.

Keywords: Himalaya clause; third parties; carriage of goods by sea; England; Canada

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

I. Introduction

This article provides an analysis of the development of Himalaya clauses in contracts of carriage of goods by sea in England and Canada. It critically examines the effectiveness of the legal bases on which Himalaya clauses have been enforced in these two jurisdictions.

Before proceeding with the analysis, it is important to outline the shipping context. An agreement for the transportation of goods by sea ("contract of carriage"), commonly evidenced by a bill of lading,[1] is concluded between the party who supplies the goods to be transported ("shipper" or "cargo owner") and the transporter ("carrier") in exchange for a reward ("freight").[2] The person who takes delivery of the goods after they have been transported is the consignee.[3] In order to perform its obligations under the contract of carriage, the carrier engages servants, agents or independent contractors, including *inter alia* stevedores, terminal or dock operators, charterers, freight forwarders, customs brokers, truckers, and so on ("third parties" or "subcontractors").

Practically, upon arrival of the ship carrying the cargo at port, the carrier enters a contract with stevedores or terminal operators to load, unload, handle or store the cargo.[4] Stevedores and terminal operators are generally considered to be independent contractors of the carrier and bailees of the cargo, thus they owe a duty to take reasonable care of it.[5] If the cargo is lost or damaged during their operations, the shipper or consignee may sue the carrier and/or the stevedores and/or the terminal operators.[6]

The contract of carriage will normally contain a Himalaya clause. A Himalaya clause, named after the ship in the case of *Adler v Dickson*,[7] is a provision in a contract of carriage that extends the carrier's exemptions or limitations of liability under the contract to third parties engaged by the carrier for the performance of the contract. In effect, a Himalaya clause is a term in a contract between A and B by which A promises B that any exemptions from or limitations of liability available under that contract to B shall also be available for the benefit of C, who is typically an employee, agent, or subcontractor of B, for the purpose of performing all or part of B's obligations under the contract.[8]

1 See *Pyrene Co Ltd v Scindia Navigation Co Ltd* [1954] 2 QB 402; *Anticosti Shipping Co v St Amand* [1959] SCR 372; Thomas J Schoenbaum, *Admiralty and Maritime Law* (6th edn, West Academic Publishing 2019) 477.
2 Schoenbaum (n 1) 475.
3 ibid.
4 ibid 493.
5 ibid 495-496.
6 ibid 494.
7 *Adler v Dickson* [1954] EWCA Civ 3, [1955] 1 QB 158.
8 HG Beale (ed), *Chitty on Contracts* (32nd edn, Sweet & Maxwell 2017) para 18-092.

Where the shipper or consignee brings an action for loss or damage of cargo against a third party engaged by the carrier, and the contract of carriage includes a Himalaya clause, the third party may seek to enforce an exemption or limitation of liability clause available to the carrier under the contract.

Even though the question whether a person may benefit from a contract to which it is not party arose 160 years ago,[9] it remains a crucial question in shipping due to the substantial involvement of third parties in the performance of carriage. As far as third parties are responsible for carrying out the carrier's obligations under the contract of carriage, legal scholars have necessitated their legal protection.[10]

Currently, the law in England and Canada allows third parties to enforce contractual provisions to their benefit, subject to certain requirements. However, the current law is the result of an incremental change of common law, following a prolonged and complex battle between traditional common law doctrine and modern commercial reality. It is impossible to fully grasp the function of Himalaya clauses without first examining their development in the common law and the judicial perplexity in finding a suitable legal basis for their enforcement.

In this article, I explore the historic development of Himalaya clauses in England and Canada. In particular, I critically analyze the legal bases for the enforcement of Himalaya clauses in the context of carriage of goods by sea, as set out by English and Canadian courts. In Section II.A, I examine the agency basis and its equivocal applicability to the relationship between shipper, carrier and third parties. In Section II.B, I examine the unilateral contract basis and its artificiality in the shipping context. In Section II.C, I discuss the exemption clause basis and the interpretation of Himalaya clauses in a way that accommodates commercial expediency. Finally, in Section II.D, I analyze the ultimate tests for the enforcement of Himalaya clauses and whether they adequately reflect the needs of modern multimodal transport. Based on the analysis, I draw certain conclusions about the overall development of Himalaya clauses in England and Canada and I argue that it is doubtful that Himalaya clauses have found their permanent home in English and Canadian law.

[9] *Tweddle v Atkinson* [1861] EWHC QB J57, (1861) 121 ER 762; *Burris v Rhind* (1899) 29 SCR 498.

[10] Bruno Zeller and Gabriël Moens, *The Himalaya Clause* (Connor Court Publishing 2020); Carlo Corcione, *Third Party Protection in Shipping* (Informa Law Routledge 2020); Donal Nolan, 'Reforming the Privity of Contract Doctrine' in Mads Andenas and Nils Jareborg (eds), *Anglo-Swedish Studies in Law* (Iustus Förlag 1999) 288.

II. The Development of the Himalaya Clause

Himalaya clauses have always sat uncomfortably with the traditional common law rule of privity of contract. The privity rule entails that no one except for the parties to a contract can be entitled under it or bound by it.[11] The privity rule has two aspects: first, a contract cannot confer enforceable rights or benefits on third parties (the "third party rule"), and second, a contract cannot impose burdens or obligations upon third parties.[12] Although the second aspect of the privity rule is relatively uncontroversial, the third party rule has been extensively criticized.[13] The third party rule provides that only the parties between whom a contractual offer and acceptance is made (the contracting parties) can enforce the resulting contract.[14] In effect, a person who is not party to a contract has no right to enforce it, even if the contract purports to confer a benefit on such person.[15] In the shipping context, the third party rule prevents third parties from enforcing an exemption or limitation clause in the contract of carriage against the claims of the shipper, even if the clause extends its protection to them.

The first case that dealt with this matter in the context of carriage (of passengers) by sea was *Adler v Dickson*.[16] The facts were simple: Adler went for a cruise on a ship and, while walking across the gangway to the ship, the gangway came adrift from the gantry at the shore and consequently Adler fell to the wharf and suffered severe injuries.[17] The Court of Appeal held that the master and boatswain of the ship were liable for negligence and could not benefit from the exemption clause in the passenger ticket, which evidenced the contract between passenger and shipowner, because it could not be construed that the clause was made for their benefit.[18] However, Lord Denning stated that, in the carriage of passengers and goods, a third party may be effectively protected under an exemption clause in the contract of carriage provided that the carrier had stipulated for the exemption clause to include the third party either by express

[11] Jack Beatson, Andrew Burrows and John Cartwright, *Anson's Law of Contract* (31st edn, OUP 2020) 613; Michael Furmston, *Law of Contract* (17th edn, OUP 2017) 556.

[12] Beale (n 8) para 18-003; John N Adams and Roger Brownsword, *Key Issues in Contract* (Reed Elsevier 1995) 125; Law Commission, *Privity of Contract: Contracts for the Benefit of Third Parties* (Law Com No 242, 1996) (Law Commission Report), para 2.1.

[13] Nolan (n 10) 288.

[14] Stephen A Smith, 'Contracts for the Benefit of Third Parties: In Defence of the Third-Party Rule' (1997) 17 OJLS 643, 644.

[15] *Dunlop Pneumatic Tyre Co Ltd v Selfridge & Co Ltd* [1915] UKHL 1, [1915] AC 847; *Beswick v Beswick* [1967] UKHL 2, [1968] AC 58; *Van Hemelryck v New Westminster Construction and Engineering Co* [1920] BCJ No 5; Guenter Treitel, 'The Battle over Privity' in Guenter Treitel (ed), *Some Landmarks of Twentieth Century Contact Law* (OUP 2002) 47, 66.

[16] *Adler* (n 7).

[17] ibid [1].

[18] ibid [14].

words or by necessary implication, and the other contracting party had assented to the exemption.[19] The reasoning of Lord Denning in *Adler* constitutes the birth of Himalaya clauses.

Currently, English and Canadian law has developed in relation to the third party rule so that a third party may avail itself of a Himalaya clause in a contract of carriage. But it took a lot of effort and imagination to find a way which, consistently with the rules of privity and consideration, could allow third parties to benefit from the contract of carriage.[20] As discussed below, the enforcement of Himalaya clauses could not easily fit into existing legal principles to avoid the third party rule and consequently English and Canadian courts struggled to find a suitable legal basis for their enforcement.

Depending on the claims of parties, Himalaya clauses have been enforced on contractual and non-contractual legal bases: contractual legal bases include agency, unilateral contracts or exemption clauses, and non-contractual legal bases include the bailment[21] or voluntary assumption of risk[22]. This article examines only the contractual legal bases.

A. Agency

In England, agency was raised as a legal basis for the enforcement of a Himalaya clause in a contract of carriage in *Scruttons Ltd v Midland Silicones*.[23] The issue in this case was whether the stevedores, who negligently damaged the cargo during loading, could rely on the limitation clause in the bill of lading between the cargo owner and the carrier and/or from the exemption clause in the stevedoring contract between them and the carrier, against the claims of the cargo owner in tort.

The stevedores argued that they could benefit from the limitation clause in the bill of lading on three grounds: (a) the word "carrier" in the limitation clause included stevedores: Viscount Simonds said that the word "carrier" should not be

[19] ibid [11].

[20] Martin Dockray, *Cases and Materials on the Carriage of Goods By Sea* (5th edn, Taylor & Francis 2004) 134.

[21] *Elder Dempster & Co Ltd and Ors v Paterson Zochonis & Co Ltd* [1924] AC 522; *The owners of cargo lately laden on board the ship or vessel "K.H. Enterprise" v. The owners of ship or vessel "Pioneer Container" Co (Hong Kong)* [1994] UKPC 9.

[22] *Scruttons Ltd v Midland Silicones* [1961] UKHL 4, [1962] AC 446 (Lord Denning); *Wilson v Darling Island Stevedoring and Lighterage Co Ltd* [1957] 95 CLR 43; For the reasoning that an exemption clause in a contract may eliminate a duty of care in tort (the "duty approach" or "tort approach"), see *Norwich City Council v Harvey* [1989] 1 WLR 828; *Pacific Associates Inc v Baxter* [1990] 1 QB 993; Philip H Clarke, 'The Reception of the Eurymedon Decision in Australia, Canada and New Zealand' (1980) 291 ICLQ 132, 148; John G Fleming, 'Employee's Tort in a Contractual Matrix: New Approaches in Canada' (1993) 13 OJLS 430, 433-434.

[23] *Scruttons* (n 22).

interpreted expansively but rather in the ordinary use of language, according to which a stevedore is not a carrier; (b) the carrier contracted as their agent: Viscount Simonds rejected this ground because the carrier engaged the stevedores as independent contractors and there was no express contractual indication that the carrier was acting as their agent; and (c) there was an implied contract between them and the cargo owner: Viscount Simonds also rejected this ground because the cargo owner did not know and it was not its concern to know about the relationship between carrier and stevedores.[24] Consequently, a contract between two unconnected parties could not be implied.

Lord Reid added that an implied contract between stevedores and cargo owner was not necessary for business efficiency in the circumstances.[25] Furthermore, Lord Reid found that the exemption clause in the stevedoring contract may protect the stevedores from liability against the carrier, but it certainly cannot bind a third party, therefore the stevedores could not enforce the exemption clause against the cargo owner.[26] The House of Lords held the stevedores liable in tort.

However, Lord Reid left open the possibility that a third party may enforce a contractual provision if the following requirements (the "agency test") are met:

> [I]f (first) the Bill of Lading makes it clear that the stevedore is intended to be protected by the provisions in it which limit liability, (secondly) the Bill of Lading makes it clear that the carrier, in addition to contracting for these provisions on his own behalf, is also contracting as agent for the stevedore that these provisions should apply to the stevedore, (thirdly) the carrier has authority from the stevedore to do that, or perhaps later ratification by the stevedore would suffice, and (fourthly) that any difficulties about consideration moving from the stevedore were overcome.[27]

This decision was later adopted by the Supreme Court of Canada in *Canadian General Electric Co v Pickford & Black Ltd*,[28] where the stevedores were found liable for negligent stowage of heavy electrical equipment onto the ship.[29] The stevedores argued that they were entitled to rely on the limitation of damages clause in the contract.[30] However, Ritchie J rejected this argument on the basis

[24] ibid 1.
[25] ibid 6.
[26] ibid 6.
[27] ibid 5-6.
[28] *Canadian General Electric Co v Pickford & Black Ltd* [1971] SCR 41.
[29] ibid 41, 45.
[30] ibid 43.

that, as discussed in *Scruttons,* the stevedores were complete strangers to the contract and should therefore accept the normal consequences of their negligence in tort.[31] The Court in *Canadian General Electric* did not apply the agency test.

Similarly, the Supreme Court of British Columbia in *Calkins & Burke Ltd v Far Eastern Steamship Co et al*[32] maintained the privity rule by reference to *Canadian General Electric* and *Scruttons*.[33] Here, the stevedores were held liable for negligence because they failed to provide in-shed storage for 37 bundles of steel tubing, as per the instructions of the consignee, and as a result the tubing was damaged by rain.[34] The stevedores argued that they could benefit from the exemption clause in the bill of lading, which extended its protection to *inter alia* any stevedores employed for the performance of any of the carrier's obligations under the bill of lading.[35] The Court rejected this argument since the exemption clause referred to the "carrier's obligations", whereas the damage of the goods occurred during storage when the goods were under the stevedores' responsibility.[36] Moreover, in applying the agency test, the Court ruled that the carrier did not have the authority, on the evidence presented, to contract as agent of the stevedores and therefore the third requirement of the agency test was not met.[37]

(i) Discussion

An agency relationship arises where a principal A authorizes an agent B to contract on his behalf with C.[38] For the purposes of the second and third requirement of the agency test, an agency relationship arises where a stevedore (principal) authorizes the carrier (agent) to enter a contract with the shipper on its behalf and stipulate an exemption clause for its benefit. In such circumstances, the stevedore may be able to rely on the exemption clause if the carrier was contracting as an agent to make the stevedore part of the contract or at least part of the exemption.[39] Agency recognizes that the agent (carrier) has the dual status of being a contracting party and an agent whose principal (stevedore) acquires the right of enforcement.[40] Thus, agency may be viewed as an exception to privity in that the

31 ibid 43, 44.
32 *Calkins & Burke Ltd v Far Eastern Steamship Co et al* [1976] BCJ No 1374 (The Suleyman Stalskiy).
33 ibid [32], [35].
34 ibid [9], [10], [14], [27].
35 ibid [28].
36 ibid [4], [48].
37 ibid [44].
38 *Halsbury's Laws of England* (5th edn, 2019) vol 22, para 136 and cited cases.
39 ibid para 199; Beale (n 8) para 15-050.
40 MH Ogilvie, 'Re-Defining Privity of Contract: Brown v. Belleville (City)' (2015) 52 Alta L Rev 731, 738.

principal (stevedore), albeit a third party to the contract concluded by its agent (carrier) and another party (shipper), is able to sue and be sued on it.[41]

Viscount Simonds in *Scruttons* explained that, since the privity rule prevents the formation of any legal principle entitling third parties to enforce contractual terms, stevedores may exclude their liability for negligence only if there is a contract between them and the cargo owner including an exemption clause to this effect.[42] To establish a contract between stevedores and cargo owner, the stevedores need to prove that the carrier was contracting with the cargo owner as agent for the stevedore or as agent for the cargo owner.[43] These agency relationships cannot be assumed lightly.

As a general rule, agency is construed only where there is a genuine intention of the parties to create an agency relationship.[44] Lord Reid stated that even if one could spell out of the bill of lading an intention to benefit the stevedore, there would still be no indication that the carrier was contracting as agent for the stevedore.[45] Thus, it is required that the parties have made clear in their contract that their relationship or the relationship between carrier and third party is one of agency.

If there is nothing in the contract indicating that the carrier is acting as agent for a subcontractor or for the shipper, an agency relationship cannot be implied in the circumstances. In shipping practice, it is uncommon for carriers to act as agents of stevedores. The main reason is because the relationship between carrier and stevedore is one of independent contractors.[46] It has been accepted

41 Law Commission Report (n 12) para 2.15; See also John D McCamus, 'Loosening the Privity Fetters: Should Common Law Canada Recognize Contracts for the Benefit of Third Parties?' (2001) 35 CBLJ 173, 181; Lance Finch and Karen Horsman, 'London Drugs LTD. v. Kuehne & Nagel International LTD' (1993) 51 Advocate (Vancouver) 409, 410.

42 *Scruttons* (n 22) 4.

43 Beatson, Burrows and Cartwright (n 11) 648.

44 John D McCamus, *The Law of Contracts* (Irwin Law 2005) 303; Brian Coote, 'Pity the Poor Stevedore!' (1981) 40 CLJ 13, 13; However, see *Borvigilant, owners of the Ship v Romina G, owners of the Ship* [2003] EWCA Civ [17] ("Lord Reid [in *Scruttons*] did not say that the clause must expressly state that the carrier is contracting as agent for the stevedores. He simply said that the contract must make that clear… whether it does so or not is a question of the construction of the… contract as a whole").

45 *Scruttons* (n 22) 6. However, see *Borvigilant* (n 44) [23]-[26] where the Court of Appeal ruled that where a contract expressly provides that it is made for the benefit of another person, there is a strong pointer to the conclusion that the contract was made on behalf of that person and that this conclusion is the only one that makes commercial sense. The conclusion in *Borvigilant* may be inconsistent with the reasoning of Lord Reid in *Scruttons* nor with the traditional rules of agency.

46 *Scruttons* (n 22) 1, 4; The Suleyman Stalskiy (n 32) [3], [7], [15]; Fleming (n 22) 430; Schoenbaum (n 1) 495-496; Corcione (n 10) para 2.32. However, third parties may be deemed to be acting as agents of the carrier on the facts, see *ITO - International Terminal Operators Ltd v Miida Electronics Inc* [1986] 1 SCR 752 [44]; *The New Zealand Shipping Co Ltd v A M Satterthwaite & Co Ltd (New Zealand)* [1974] UKPC 4, [1975] AC 154 (*The Eurymedon*) 1, 4; Corcione (n 10) para 3.63 (freight forwarders can act as agents of the carrier in issuing the bill of lading); Chester D Hooper, 'Legal Relationships:

that the role of independent contractors in the chain of carriage is distinct from the role of employees or agents, and there is no strong justification for why independent contractors must be protected under the contract of carriage.[47] It has been argued that the mere fact that independent contractors perform services for the carrier does not justify their entitlement to enforce an exemption or limitation clause in the contract.[48] This position is reinforced by the wording of Article IV(2) of the Hague-Visby Rules[49] which provides that the carrier's servants or agents (but not independent contractors) may have the benefit of the carrier's exemptions and limitations of liability under the Rules. Thus, the distinction between employees or agents and independent contractors became more acute. While employees and agents are engaged to perform a contract *of* service, independent contractors are engaged to perform a contract *for* service.[50] In other words, contrary to employees or agents, contractors do not provide the carrier's services to others but instead they provide independent and professional services to others for the carrier in exchange for a reward. Particularly, contrary to employees, stevedores assume responsibility for performing the carriage and for insuring against the risk of loss or damage of cargo, and they acquire the necessary experience and knowledge to contract for indemnity or protection against liability.[51] It follows that, as long as the carrier and stevedore are independent contractors, there is no reason to *assume* that the carrier had contracted for the stevedore.[52]

Similarly, it is uncommon for carriers to act as agents of shippers. The purpose of the contract of carriage is to allocate the risks of transport and insurance and not to authorize the carrier to contract on behalf of the shipper. This is reinforced by the fact that the shipper is not bound by any downstream contract between the carrier and its subcontractors.[53] Thus, a contract of carriage cannot imply that the carrier engages subcontractors as agent of the shipper, unless there is an express stipulation to that effect.

Terminal Owners, Operators, and Users' (1990) 64 Tul L Rev 595, 596-597 (terminal operators can act as agents of the carrier in performing its obligations under the bill of lading).

47 Corcione (n 10) para 2.39.

48 Samir Mankabady, 'Rights and Immunities of the Carrier's Servants or Agents' (1973) 5 J Mar L & Com 111, 113.

49 Protocol to Amend the International Convention for the Unification of Certain Rules of Law Relating to Bills of Lading 1968 (The Hague-Visby Rules), which has been enacted in England by the Carriage of Goods by Sea Act 1971 and in Canada by the Marine Liability Act (SC 2001, c 6).

50 This distinction may be drawn by the "control" test, the "organization" test, the "representative" test or the test examining the nature and totality of the relationship between the parties, see Jonathan Burnett, 'Avoiding Difficult Questions: Vicarious Liability and Independent Contractors in Sweeney v Boylan Nominees' (2007) 29 Syd L Rev 163, 166, 169.

51 Fleming (n 22) 430.

52 *Scruttons* (n 22) 1, 4.

53 *Scruttons* (n 22) 6; Treitel (n 15) 59, 64.

It is clear from the decisions in *Scruttons* and *The Suleyman Stalskiy* that an agency relationship between carrier and shipper or between carrier and subcontractor cannot be implied, even if the contract expressly conferred a benefit on the subcontractor. Hence, third parties cannot escape the privity rule by *implied* agency, but they can do so if the contracting parties have used expressed agency as a means of conferring on them a right of enforcement.[54]

On the one hand, the requirement of expressed agency is arguably prejudicial to the interests of those third parties who are not in a position to influence the drafting of the principal contract. Although some stevedoring contracts may obligate the carrier as agent to include an exemption clause in the principal contract to their benefit,[55] most of the downstream contracts between carrier and subcontractors or between subcontractors and sub-subcontractors will be concluded after the bill of lading is issued and therefore most of the downstream parties will not be able to stipulate for expressed agency. It follows that the second and third requirements of the agency test are difficult to meet.

On the other hand, however, the requirement of expressed agency ensures that the shipper *knows* at the time of contract whether the carrier will engage third parties as an agent. Viscount Simonds noted that it is difficult to accept that the carrier was contracting as agent for an undisclosed principal where this is not expressly stated in the contract.[56] In agency for undisclosed principals, the principal has the right to enforce a contract made on his behalf by an agent, even though the agent was not known to be acting for a principal.[57] This is inconsistent with the privity rule since the third party (the shipper) may find itself in a contractual relationship with someone of whom it has never heard, and with whom he never intended to contract.[58] When the contract of carriage is made, the contracting parties are usually unaware of the existence, identity, or number of third parties that need to be engaged for the full performance of the carriage; any downstream contracts between carrier and subcontractors may be concluded after the bill of lading is issued, and the chain of downstream contracts may continue to expand even after the goods have been shipped. Can the agency of the carrier be assumed for any undisclosed subcontractor? Even if it was assumed that the carrier was acting as agent for one undisclosed principal, for example a terminal operator, it could hardly be assumed that this agency relationship expands to an indeterminate number of third parties subsequently engaged by the terminal operator. For example, where a stevedore contracts with the carrier through a terminal agent, it is questionable whether the third requirement of the agency test

[54] Robert E Forbes, 'Practical Approaches to Privity of Contract Problems' (2002) 37 CBLJ 357, 361.
[55] See for example *ITO* (n 46) [4].
[56] *Scruttons* (n 22) 1, 4.
[57] Patrick S Atiyah and Stephen A Smith, *Introduction to the Law of Contract* (6th edn, OUP 2006) 346.
[58] ibid.

applies to the stevedore who is the subcontractor of the subcontractor (sub-subcontractor).[59] If sub-subcontractors could benefit from the agency of the carrier, the shipper could be exposed to an indeterminate risk of suffering loss or damage of cargo from the negligence of an indeterminate number of third parties who could benefit from an exemption clause in the bill of lading. For this reason, it has been accepted that not every potential or intended third party beneficiary should be permitted to enforce the contract; there must be a limit on the right of third parties to sue.[60]

On the same grounds, it is questionable whether a third party who has not been identified at the time of contract can ratify at all or whether it can ratify without the consent of the other contracting party.[61] It has been argued that where the contract between carrier and subcontractor expressly stipulates that the execution of the contract by the subcontractor shall be deemed to be ratification of the exemption clause in the bill of lading between shipper and carrier, it will likely amount to sufficient ratification.[62] But it is still required to include such intention in the contract in an express manner.

It is thus extremely important for the contracting parties to disclose at the time of contract, even in the abstract, which third parties will benefit from the exemption clause so as to properly arrange insurance and allocate the risks.[63] Whether third parties can benefit from the exemption clause is a key factor in setting the freight rate and insurance.[64] If the carrier obtains a broad exemption clause for itself and its subcontractors, the freight rate will be lower; a shipper who paid a more affordable freight will be generally prevented from evading the exemption clause by suing the carrier's subcontractors.[65] If the carrier bears a higher risk of loss or damage of cargo during transit, the freight rate will be higher.[66] The greater the liability of the carrier and its subcontractors, the greater the rate of freight. It derives that if the shipper does not know at the time of contract which and how many third parties will be exempt from liability, it cannot

59 William Tetley, 'The Himalaya Clause, Stipulation pour Autrui Non-Responsibility Clauses and Gross Negligence under the Civil Code' (1979) 20 C de D 449, 453.

60 Nolan (n 10) 299.

61 Forbes (n 54) 364, note 16.

62 ibid.

63 *Valmet Paper Machinery Inc v Hapag-Lloyd AG* [2002] BCJ No 1271 [53] ("It is desirable that parties be able to ascertain, prior to shipping cargo, where the risk will lie, so that insurance can be placed appropriately and without duplication").

64 C A Ying, 'The Himalaya Clause Revisited' (1980) 22 Malaya L Rev 212, 221; Corcione (n 10) para 2.28; William T J De La Mare, 'Jurisprudential Problems of Attribution of Liability in the Area of Admiralty Contracts for Carriage following Norfolk Southern Railway v. Kirby' (2006) 22 Conn J Intl L 203, 209.

65 Corcione (n 10) para 9.91-9.92.

66 De La Mare (n 64) 210.

negotiate the freight rate and the scope of the exemption clause for its best interests.

For all of the foregoing reasons, implied agency is a weak legal basis for the enforcement of Himalaya clauses. The artificial extension of the agency exception is a precarious device for avoiding the third party rule.[67] In any case, the agency test "did not open the door very wide"[68]. In fact, it has been argued that stevedores will rarely succeed in proving the requirements of the agency test.[69]

The technicalities arising out of the third requirement of the agency test were eventually abolished by both English and Canadian courts and therefore third parties are no longer required to prove authority or ratification in order to benefit from a Himalaya clause.[70]

B. Unilateral Contract

The second legal basis for the enforcement of Himalaya clauses is the finding of an offer for a unilateral contract from the shipper to a third party through the agency of the carrier. This basis was introduced by Lord Wilberforce in *The Eurymedon*[71] stating that:

> [T]he Bill of Lading brought into existence a bargain initially unilateral but capable of becoming mutual, between the shippers and the [stevedore], made through the carrier as agent. This became a full contract when the [stevedore] performed services by discharging the goods. The performance of these services for the benefit of the shipper was the consideration for the agreement by the shipper that the [stevedore] should have the benefit of the exemptions and limitations contained in the Bill of Lading.[72]

A unilateral contract is deemed to be an open offer that matures into a contract when accepted by performance; in the shipping context, the shipper's promise to exempt the carrier's subcontractors from liability constitutes an offer to be accepted by them by performance.[73]

[67] McCamus, *The Law of Contracts* (n 44) 303.
[68] William Tetley, 'Himalaya Clause - Heresy or Genius' (1977) 9 J Mar L & Com 111, 121.
[69] Coote (n 44) 13.
[70] *Valmet* (n 63) [45], [51]-[54]; *Canada Maritime Ltd v Oerlikon Aerospace Inc* [1998] EWCA Civ 170 [23] (Hirst LJ held that there is no need to consider technicalities such as ratification since the privity rule "may well be tottering on the brink of collapse" in the carriage of goods by sea).
[71] *The Eurymedon* (n 46).
[72] ibid 4.
[73] ibid 5; See also *Carlill v Carbolic Smoke Ball Co* [1892] EWCA Civ 1, [1893] I QB 256.

In *The Eurymedon*, the issue was whether the stevedores, who negligently damaged the cargo while unloading, could benefit from the exemption clause in the bill of lading between carrier and cargo owner. The bill of lading included a Himalaya clause which expressly protected the carrier's servants, agents and independent contractors from any liability against the cargo owner.[74]

Lord Wilberforce ruled that the four requirements of the agency test were met: (a) the carrier had clearly stipulated for certain exemptions in the bill of lading for the benefit of third parties engaged by it;[75] (b) the carrier was contracting as agent of the stevedores since, for many years, the stevedores carried out the stevedoring operations of the vessels of the carrier, who was a wholly owned subsidiary of the stevedores;[76] (c) the carrier was authorized by the stevedores to contract as their agent;[77] and (d) the consideration moving from the stevedores to the cargo owner was the actual performance of the stevedoring services as acceptance to the offer for a unilateral contract.[78] Hence, Lord Wilberforce concluded that the stevedores could benefit from the exemption clause in the bill of lading so as to give effect to the clear intentions of the parties.[79]

This reasoning was adopted by the Supreme Court of Canada in *ITO*[80] where 169 cartons of calculators were stolen from the premises of the stevedores (ITO) who were engaged by the carrier (Mitsui) to unload and store the calculators until delivery to the consignee (Miida).[81] The bill of lading between Mitsui and Miida exempted Mitsui's liability for loss of cargo and extended this exemption to *inter alia* stevedores via a Himalaya clause.[82] In addition, clause 7 of the stevedoring contract between Mitsui and ITO provided that Mitsui would include ITO as an express beneficiary of all rights and exemptions under any bill of lading issued.[83] The issue was whether ITO could invoke the Himalaya clause in the bill of lading to exclude its tortious liability.

The Supreme Court ruled that in order for ITO to benefit from the bill of lading, there must be a link between it and Miida which would sufficiently bring ITO into the contractual arrangement between Mitsui and Miida.[84] Without providing any further analysis on this point, the Court proceeded to apply the agency test as set out *Scruttons* and applied in *The Eurymedon*, and found that all

74 *The Eurymedon* (n 46) 2.
75 ibid 3.
76 ibid 1, 4.
77 ibid 4.
78 ibid 4, 5.
79 ibid 6.
80 *ITO* (n 46).
81 ibid [2].
82 ibid [4].
83 ibid.
84 ibid [43].

requirements were met: (a) the Himalaya clause explicitly covered stevedores; (b) the Himalaya clause clarified that Mitsui should be deemed to be contracting as agent for stevedores; (c) Mitsui had authority from ITO to contract on its behalf pursuant to clause 7 of the stevedoring contract; and (d) any difficulties of consideration moving from ITO to Miida are overcome by reference to the decision by Lord Wilberforce in *The Eurymedon*.[85] It was held that ITO could rely on the Himalaya clause in the bill of lading.

(i) Discussion

The decision in *The Eurymedon* has been characterized as sounding the "death knell" to the privity rule in the carriage of goods by sea.[86]

The implication of a unilateral contract raised some well-founded concerns. The first concern is that, in theory, the legal basis of unilateral contract should free stevedores from the need to prove the agency test since the unilateral contract would be formed, not at the same time as the contract of carriage, but later when the services are performed in reliance on the offer of immunity.[87] However, the Privy Council treated the Himalaya clause as an immediate bargain rather than an offer for a future contract, and thus the agency test has come to be a precondition to the formation of a unilateral contract.[88] Thus, stevedores are burdened to prove both the agency test and a unilateral contract.

The second concern is reflected in the dissenting opinion of Lord Simon in *The Eurymedon*, which stressed the difficulty in accepting that the Himalaya clause in the bill of lading established a contract or bare promise (*pactum* or *nudum pactum*) between cargo owner and stevedores that consisted only of an exemption clause.[89] Since the stevedores are under no contractual obligation against the cargo owner, the scope of the exemption clause (in the offer for unilateral contract) is unclear.[90] It is a fair point that the content of an offer for a unilateral contract containing only an exemption clause cannot be determined with certainty. It is difficult to extract the content of the relationship between cargo owner and stevedores from the Himalaya clause or the contract of carriage in general between cargo owner and carrier. Additionally, it is doubtful whether the cargo owner, by contracting with the carrier, intended to make an open offer towards and enter a direct contractual relationship with any third party engaged by the carrier for the performance of carriage. To assume that the cargo owner made an offer for a

[85] ibid [44].
[86] Ying (n 64) 214.
[87] Coote (n 44) 14.
[88] ibid.
[89] *The Eurymedon* (n 46) 15.
[90] ibid.

unilateral contract to a third party through the carrier would be to re-write the Himalaya clause.[91]

The third concern relates to consideration. The courts in *The Eurymedon* and *ITO* accepted that defective performance (damaged and lost cargo) qualified as sufficient consideration for a unilateral contract. In the traditional formation of unilateral contracts, the *full* and *proper* performance of an act specified in the offer constitutes the promisee's acceptance and consideration.[92] In this regard, the carrier must make its subcontractors aware of the provisions of the exemption clause in the contract (and in the offer), so that the subcontractors' performance of the services qualifies as acceptance and consideration corresponding to the offer.[93] In order for defective performance to fit the traditional unilateral form, it must be construed that the offer required performance in any manner, even one of reduced quality.[94] Thus, defective performance may amount to sufficient consideration for a unilateral contract only if it is construed that the offer allowed so. Once again, it is extremely difficult to determine the content of a non-existent offer for the purpose of providing corresponding consideration.

Even if it is construed that the purpose of formulating a unilateral contract is exactly to enable stevedores to exclude their liability for defective performance, a line must be drawn in cases where the performance is so defective that cannot logically amount to sufficient consideration.[95] It is unclear whether the court would reach the same conclusion if all of the cargo was stolen and as a result there was no-delivery.[96] Should the court accept that stevedores have provided sufficient consideration if they negligently lost or damaged one or fifty or one hundred packages out of a hundred-packages cargo? If the answer is yes, the principle of consideration would be applied in a distorted manner, contravening the traditional formation of unilateral contracts that require full and proper consideration.

Furthermore, if defective consideration amounted to sufficient consideration, stevedores would be under no obligation or expectation to perform satisfactorily. By accepting that stevedores provide sufficient consideration irrespective of the magnitude of the defect in their performance, stevedores are not encouraged to carry out their operations with diligence nor to promote loss prevention tactics. It is a matter of policy in the context of carriage of goods that

91 Law Commission Report (n 12) para 2.27.

92 Linda C Reif, 'A Comment on ITO Ltd. v. Miida Electronics Inc. - The Supreme Court of Canada, Privity of Contract and the Himalaya Case' (1988) 26 Alta L Rev 372, 381; Clarke (n 22) 137; *Carlill* (73); *The Queen (Ont.) v Ron Engineering* [1981] 1 SCR 111; *Daulia Ltd v Four Millbank Nominees Ltd* [1977] EWCA Civ 5, [1978] Ch 231.

93 Forbes (n 54) 363; Ying (n 64) 215-216.

94 Reif (n 92) 381.

95 ibid 384.

96 ibid.

the person who causes damage should be held liable under the law, otherwise it will continue to be negligent.[97] This policy was also considered by the High Court of Australia in ruling that a decision in favour of the cargo owner and against the stevedores would encourage carriers to insist on reasonable diligence on the part of their subcontractors.[98]

These concerns were not addressed neither in *The Eurymedon* nor *ITO*. In fact, it has been commented that the issues arising from *The Eurymedon* "were swept under the carpet" by McIntyre J in *ITO*.[99] It appears from both decisions that courts were not yet ready to form an independent and uniform exception to the privity rule allowing third parties to enforce contractual terms. Instead, they preferred to justify the enforcement of Himalaya clauses by traditional principles of contract law. McIntyre J in *ITO* explained that Himalaya clauses cannot be enforced on the basis of a "jus tertii" because this would weaken the privity rule, but they may be enforced by placing the parties' relationship into the traditional mold of contract law.[100] This reasoning is also found in the following passage by Lord Wilberforce in *The Eurymedon*, which has been also adopted by McIntyre J:

> It is only the precise analysis of this complex of relations into the classical offer and acceptance, with identifiable consideration, that seems to present difficulty, but this same difficulty exists in many situations of daily life… English law, having committed itself to a rather technical and schematic doctrine of contract, in application takes a practical approach, often at the cost of forcing the facts to fit uneasily into the marked slots of offer, acceptance and consideration.[101]

I disagree with this approach. The fact that English law has adopted a technical and schematic doctrine which proved to be problematic in other categories of cases, does not justify the application of such problematic doctrine to maritime cases. The alleged practical approach followed by courts not only forces "the facts to fit uneasily into the marked slots of offer, acceptance and consideration" but also stretches traditional principles of contract law like unilateral offer and consideration. It is argued that the purity of contractual

97 Tetley, 'Himalaya Clause - Heresy or Genius' (n 68) 122.

98 *Port Jackson Stevedoring Pty Ltd v Salmond & Spraggon (Aust) Pty Ltd* [1978] HCA 8, (1977) 139 CLR 231 overturning the decision of the Privy Council in *Port Jackson Stevedoring Pty Ltd v Salmond and Spraggon (Australia) Pty Ltd (Australia)* [1980] UKPC 23, [1981] 1 WLR 138 (*The New York Star*).

99 Reif (n 92) 382.

100 *ITO* (n 46) [37].

101 *The Eurymedon* (n 46) 4; *ITO* (n 46) [36].

principles is much more severe than protecting negligent third parties against liability.

Although the uniform enforcement of Himalaya clauses across jurisdictions is crucial, it needs to be justified on concrete foundations and proper reasoning. The decisions in *ITO* and *The Eurymedon* do not constitute concrete foundations but merely an acknowledgement that commercial practice necessitates the protection of third parties. In fact, the reasons upon which *ITO* was decided, for example the intentions of the parties, commercial reality, uniformity, legal certainty, proper allocation of risk and insurance, have been characterized as "plain badges" brought forth to achieve the enforcement of Himalaya clauses in Canadian maritime law, even if it leads to the distortion of traditional contractual principles.[102] This is further discussed below at Section II.C.(i).

Based on the foregoing, the reasoning of Lord Wilberforce in *The Eurymedon*, as adopted by *ITO*, has been criticized as artificial,[103] technical,[104] and as sliding to legal fiction[105]. Later in *The Mahkutai*,[106] Lord Goff accepted that insofar as the enforcement of Himalaya clauses lays on a unilateral contract reasoning, technical points of contract and agency will inevitably arise.[107]

C. Exemption clauses

The third basis is the construction and enforcement of exemption clauses in favor of third parties for the purpose of giving effect to the intentions of the contracting parties and to commercial reality. This basis arose from the following reasoning of Lord Wilberforce in *The Eurymedon*:

> [T]o give the [stevedore] the benefit of the exemptions and limitations contained in the Bill of Lading is to give effect to the clear intentions of a commercial document… [There is] no reason to strain the law or the facts in order to defeat these intentions. It should not be overlooked that the effect of denying validity to the

[102] Reif (n 92) 382.

[103] Law Commission Report (n 12) para 2.27; See also *Southern Water Authority v Carey* [1985] 2 All ER 1077 (the court doubted that the unilateral contract reasoning is applicable beyond the carriage of goods context as it is "uncomfortably artificial"); For the argument that the conception of this type of contracts as "bilateral contracts" extinguishes artificiality, see Robert Stevens, 'The Contracts (Rights of Third Parties) Act 1999' (2004) 120 LQR 292, 304-305.

[104] Clarke (n 22) 148, 150.

[105] Reif (n 92) 380.

[106] *The owners and/or demise charterers of the ship or vessel "Mahkutai" (Indonesian Flag) v The owners of lately laden on board the ship or vessel "Mahkutai" (Indonesian Flag) Co (Hong Kong)* [1996] UKPC 71, [1996] AC 650.

[107] ibid 11-12.

> clause would be to encourage actions against servants, agents and independent contractors in order to get round exemptions… accepted by shippers against carriers, the existence, and presumed efficacy, of which is reflected in the rates of freight.[108]

Even if a third party seeking to rely on a Himalaya clause proves that the requirements of the agency test or unilateral contract are met, it is a matter of construction whether it can rely or not. In general, courts have interpreted Himalaya or exemption clauses broadly.

The leading example in which the court interpreted an exemption clause broadly so as to give effect to the parties' intentions is *ITO*. The Supreme Court of Canada construed that the exemption clause in the bill of lading implied an exemption for liability in negligence, even though it explicitly exempted liability only "for any delay, non-delivery, misdelivery or loss of or damage to or in connection with the goods",[109] because the wording of the Himalaya clause was wide enough to include negligence within the contemplation of the parties in formulating the contract.[110] The Court noted that the implication is reinforced by the fact that the exemption clause only applied to a small part of the full agreed carriage, *i.e.* after loading and before discharge.[111]

Similarly, in *The Eurymedon*, where the Himalaya clause exempted the stevedores' liability for any "act, neglect or default" in the course of their employment,[112] the Privy Council, although acknowledging that the drafting was defective, found that the stevedores were exempted from all liability, including negligence, without providing further analysis.[113]

The leading example in which the court interpreted the contract broadly so as to give effect to commercial reality is *The New York Star*[114]. Here, the Privy Council extended the scope of application of the bill of lading to the period after discharge, contrary to the explicit contractual provisions, to protect the stevedores from liability. In this case, the loss occurred after the cargo was discharged from

[108] *The Eurymedon* (n 46) 6. See also *"Starsin", Owners of cargo & Ors v "Starsin", Owners and/or demise charterers of* [2003] UKHL 12, [2004] 1 AC 715 [57]; *Borvigilant* (n 44) [23].

[109] *ITO* (n 46) [4].

[110] ibid [48].

[111] ibid.

[112] *The Eurymedon* (n 46) 2.

[113] ibid 3. See also *The Eurymedon* (n 46) 14 (in his dissenting judgment, Lord Simon stated that the Himalaya clause should not be enforced because the wording "act, neglect or default" appears to extend to theft or even malicious damage and not tortious liability).

[114] *The New York Star* (n 98) (in this case, the cargo was stolen from a shed on the wharf under the control of the stevedores and the cargo owner brought an action against the stevedores for negligence. The bill of lading between cargo owner and carrier included a Himalaya clause which extended the carrier's immunities to its independent contractors. The Privy Council did not provide an analysis of the agency test but simply mentioned that all of the requirements were found to be met by the lower courts).

the vessel, when the stevedores were not performing the carrier's obligations under the bill of lading but rather, they were acting as bailees of the cargo[115]. Lord Wilberforce explained that, in light of the commercial practice at ports, the consignees normally take delivery of the cargo after some period of storage on the wharf and not directly from the ship's rail, as it was provided for in the bill of lading.[116] Taking this practice into consideration, he concluded that the carrier would inevitably employ a stevedore to store the cargo on the wharf, therefore the carrier's immunity extended to the period after discharge, despite the expressly defined period in the bill of lading.[117] On this basis, the stevedores were exempted from liability.

It is nevertheless interesting to note that, normally, where a bill of lading applies to the period after loading and before discharge, any loss or damage of cargo caused by a negligent stevedore *while* loading or *during* discharge will normally not be exempted from liability.[118]

Finally, the broad construction of exemption clauses is also evident by the enforcement of circular indemnity clauses embedded in Himalaya clauses.[119] A circular indemnity clause consists of (a) a promise not to sue, by which the shipper agrees not to bring any action against the carrier's subcontractors, and (b) an indemnity clause, by which the shipper agrees to indemnify the carrier if any action is brought contrary to the promise not to sue.[120] Consequently, the shipper ends up facing its own claims.[121]

(i) Discussion

Should courts construe Himalaya clauses in such a broad manner for the benefit of third parties? It is true that whether an exemption clause extends its protection to a third party is decided on ordinary principles of construction.[122] A contract of carriage is construed like any other contract, by considering its terms, the intentions of the parties,[123] all surrounding circumstances,[124] trade practices

115 ibid 5.
116 ibid 6.
117 ibid 7.
118 Ying (n 64) 216; See also *Raymond Burke Motors Ltd v Mersey Docks & Harbour Co* [1986] 1 Lloyd's Rep 155; *Southampton Cargo Handling Plc v Lotus Cars Ltd & Ors* [2000] EWCA Civ 252, [2001] CLC 25 (The Rigoletto).
119 *Nippon Yusen Kubishiki Kaisha v Golden Strait Corporation* [2003] EWHC 16 (Comm), [2003] 2 Lloyd's Rep 592; *Bombardier Inc v Canadian Pacific Ltd* [1988] OJ No 1807.
120 Corcione (n 10) para 2.54.
121 William Tetley, 'The Himalaya clause – Revisited' (2003) 9 JIML 3, 27.
122 *Halsbury's Laws of England* (5th edn, 2019) vol 22, para 188, 199; *Borvigilant* (n 44) [17]-[33].
123 Schoenbaum (n 1) 476.
124 *Nisshin Shipping Co Ltd v Cleaves & Co Ltd & Ors* [2003] EWHC 2602 (Comm), [2004] 1 Lloyd's Rep 38 [23].

and implied terms.[125] In determining what was within the reasonable contemplation of the parties, courts must consider that the bill of lading is a commercial contract that determines which party will bear insurance.[126]

Nevertheless, it is a traditional common law rule that exemption clauses should be interpreted strictly and, particularly, exemption clauses that generally exempt all liability will be construed more restrictively than limitation clauses.[127] This means that a Himalaya clause, which does not expressly include the third party seeking to rely on it or exempts such third party from all liability, will be construed strictly and therefore the third party will not be able to enforce it.

This rule has been considerably relaxed in common law. In England, the House of Lords refused to adopt a rule by which courts may invalidate or deprive the effect of exemption clauses, even in the event of fundamental breach.[128] Particularly, the House of Lords explained that the Parliament has passed the Unfair Contract Terms Act 1977 (UCTA 1977) which applies to consumer contracts or standard form contracts and enables the application of exemption clauses only if they are just and reasonable, but the Parliament has nevertheless refrained from regulating commercial contracts, where the parties are not of unequal bargaining power and their risks are normally borne by insurance.[129] This shows that there is no need for judicial intervention to exemption clauses in commercial contracts and the parties are free to allocate the risks as they think appropriate.[130]

This approach has been adopted by the Supreme Court of Canada.[131] Since there is no similar legislation with UCTA 1977 in Canada, the Supreme Court ruled that Canadian courts are responsible for deciding whether an exemption clause is enforceable or not in the circumstances for the purpose of giving effect to the bargain of the contracting parties.[132] In this analysis, exemption clauses should be given their natural and true construction so that their meaning and effect is fully understood as the contracting parties agreed to at the time of contract.[133]

[125] Beatson, Burrows and Cartwright (n 11) 626-627.

[126] *ITO* (n 46) [48].

[127] Reif (n 92) 386 and cited cases.

[128] *Suisse Atlantique v Rotterdamsche Kolen Centrale* [1967] AC 361; See *Photo Production Ltd v Securicor Transport Ltd* [1980] UKHL 2, [1980] AC 827 where the House of Lords overruled the traditional doctrine by which a breaching party could not benefit from the exemption clause in the contract if there was a fundamental breach (for the traditional doctrine, see *Harbutt's "Plasticine" Ltd v Wayne Tank & Pump Co Ltd* [1970] 1 QB 447; *Karsales (Harrow) Ltd v Wallis* [1956] EWCA Civ 4, [1956] 1 WLR 936).

[129] *Photo Production* (n 128) 3-4.

[130] ibid 4; *The Starsin* (n 108) [57].

[131] *Hunter Engineering Co v Syncrude Canada Ltd* [1989] 1 SCR 426; *Beaufort Realties (1964) Inc v Belcourt Construction (Ottawa) Ltd* [1980] 2 SCR 718.

[132] *Hunter* (n 131) [153].

[133] ibid [153].

Exemption clauses cannot be considered in isolation from the other contractual provisions and the circumstances in which they were entered into.[134]

This rationale was applied by the Supreme Court in *ITO* in concluding that, in commercial transactions where the parties have equal bargaining power, exemption clauses are likely to be interpreted broadly, considering that the parties have already allocated their risks by insurance.[135] Although the Court acknowledged that a general exemption from all liability in a contract will not *per se* exclude negligence,[136] McIntyre J explained that an exemption clause which does not explicitly exclude liability for negligence, but its wording is so wide and clear that it could do so implicitly, may be interpreted as excluding negligence unless the parties intended otherwise.[137] In construing the parties' intentions in relation to the scope of the exemption clause, the Court must consider whether negligence was the only possible type of liability upon which the clause could operate in the circumstances, as well as whether the parties could be deemed to have contemplated such possibility.[138] Since ITO was a sub-bailee of the cargo after it was unloaded and the bill of lading absolved the carrier's liability for the period before loading or after discharge, the Court found that the only reasonable head of ITO's liability for loss of cargo was its failure to take reasonable care of the cargo – negligence.[139]

In my view, where the contracting parties are sophisticated entities, for example well-established corporations with equal bargaining power, extended insurance coverages and easy access to legal advice, the standard of contract drafting should be raised, at least in relation to third party protection. Where the contracting parties are professional or experienced actors in the shipping industry, it should be reasonably expected that they have expressly included their bargain in the contract. Anything that is not expressly provided for in the contract should be presumed to be outside the contemplation of the parties at the time of contract. Especially in relation to exemption clauses, which are usually subject to long negotiations and careful drafting, the actual wording of the clause is key in determining the parties' intentions and therefore should be given priority over commercial practices.

The implication of words or meanings to exemption clauses for the purpose of exempting the liability of third parties may lead to considerable re-drafting of the clause by courts. Lord Simon, in his dissenting opinion in *The Eurymedon*, explained that as long as the contract did not expressly exclude the stevedores'

[134] ibid [151].
[135] Reif (n 92) 386.
[136] *ITO* (n 46) [48].
[137] ibid 45.
[138] Reif (n 92) 385-386.
[139] See *ITO* (n 46) [48].

tortious liability, to imply it would be to re-write it.[140] This type of judicial intervention to the contract should always be balanced with the two original criteria set out by Lord Denning in *Adler*: whether it is necessary for business efficiency to imply the third party from the exemption clause and whether the injured party had assented to the exemption.[141] The courts in *The New York Star* and *ITO* did not provide an analysis on these two criteria; they were too quick in allowing negligent third parties to benefit from a broadly interpreted exemption clause. However, where a third party is found to be negligent for loss of cargo and there is no contractual indication that the parties intended to protect it against liability for negligence, it is not necessary for business efficiency to extend the scope of the Himalaya clause by implication; the contract operates properly without this implication and the third party will be liable for the normal consequences of its tort. By extending the scope of the Himalaya clause by implication, the court may interfere with the actual intentions of the parties rather that give effect thereto.

This concern was also expressed by Fullagar J in *Wilson v Darling Island Stevedoring and Lighterage Co*:[142]

> The common law has… allowed the validity of provisions of a contract which limit or exclude liability for negligence. But it has always frowned on such provisions and insisted on construing them strictly… And yet we seem to discern… a curious, and seemingly irresistible, anxiety to save grossly negligent people from the normal consequences of their negligence.[143]

The dissent in *The Eurymedon* adopted the foregoing passage and ruled that the anxiety to protect negligent parties cannot give unnatural or artificial meaning to exemption clauses which, if construed strictly, do not exempt liability in negligence.[144]

Finally, a central judicial concern in such cases is that it is commercially undesirable to allow shippers to circumvent exemption clauses by suing the carrier's subcontractors in tort because it would undermine the general purpose of exemption clauses and would redistribute the risk between the parties, as reflected in the rate of freight and insurance.[145] Put simply, the non-enforcement of Himalaya clauses would encourage shippers to bring actions against the carrier's

[140] *The Eurymedon* (n 46) 15.
[141] *Adler* (n 17) [11].
[142] *Wilson v Darling Island Stevedoring and Lighterage Co* [1957] 95 CLR 43.
[143] ibid [10].
[144] *The Eurymedon* (n 46) 9.
[145] *The Mahkutai* (n 106) 8; Beatson, Burrows and Cartwright (n 11) 651; Timothy Liau, 'Privity: Rights, Standing, and the Road Not Taken' (2021) 41 OJLS 803, 815-816; Coote (n 44) 15.

subcontractors to get round exemptions.[146] Nevertheless, this concern is not supported. It is equally commercially undesirable to allow any number of unidentified third parties to avoid the consequences of their negligence by relying on the carrier's exemptions and limitations. The wide enforcement of Himalaya clauses merely benefits the interests of great fleet-owning nations whose ocean carriers and their subcontractors are protected from liability; it is nonetheless detrimental to the interests of those countries, such as Australia, which largely depends on those fleets for import and export trade.[147] Besides, if a carrier wishes to prevent a shipper from suing its subcontractors in tort, the carrier may either explicitly exclude the tortious liability of its subcontractors or include a circular indemnity clause in the Himalaya clause. If the carrier does not do so, and there is no other indication in the contract as whole that third parties are protected from tortious liability, it is presumed that the parties did not intend to protect third parties *to that extent* and thus this intention should be given effect by not allowing the third party to enforce the Himalaya clause.

Overall, the broad construction of exemption clauses for the purpose of giving effect to the parties' intentions and commercial reality is not a convincing legal basis. The "commercial reality" and "intentions of the parties" rationale has been criticized as inadequate to justify by itself an exception to the privity rule in favor of third parties.[148] Additionally, it has been submitted that it is difficult to justify ad hoc contractual rights held by third parties on the basis of "commercial necessity" because it requires the artificial creation of an agreement that does not exist.[149] The same is true for the basis of the "intentions of the parties" which requires the meeting of the parties' minds through external communication of offer, acceptance, and consideration.[150] Thus, the deficient justification of contractual rights of third parties for purposes of practicality may result in normative incoherence in contract law.[151]

D. Generic test

It appears that both English and Canadian courts have been generally willing to enforce Himalaya clauses in favour of third parties. The Privy Council emphasized that there is great readiness to accept the doctrine of vicarious immunity for pragmatic and commercial reasons and that, on the appropriate

146 *The Eurymedon* (n 46) 6.
147 *Port Jackson Stevedoring* (n 98) [13].
148 Jason W Neyers, 'Explaining the Principled Exception to Privity of Contract' (2007) 52 McGill LJ 757.
149 Liau (n 145) 829.
150 ibid.
151 ibid.

facts, courts may establish "a fully-fledged exception" to privity "thus escaping from all the technicalities with which courts are now faced in English law"[152].

However, as discussed in the analysis, the legal bases for Himalaya clauses have been extensively criticized by judges and scholars. In response to this criticism, English and Canadian law have adopted a generic test as an ultimate solution to the issue of third-party protection: England established a statutory test, whereas Canada established a common law test to allow third parties to enforce contractual terms. The word "generic" is used as both tests are not specifically designed for contracts for the carriage of goods by sea, but they apply to all types of contract (although the English test is subject to exceptions).

(i) The English Test

England has enacted the Contracts (Rights of Third Parties) Act 1999 (1999 Act), which largely implemented the recommendations of the Law Commission's Report.[153] The 1999 Act applies to all contracts except for those listed in section 6. Particularly, section 6(5) excludes the application of the 1999 Act to contracts for the carriage of goods by sea, rail, road, and air, except that third parties may take advantage of exclusion or limitation clauses thereunder. The explanatory notes of the 1999 Act clarify that section 6(5) enables a third party to enforce an exemption or limitation clause if such clause extends its protection to servants, agents and independent contractors engaged by the carrier in the loading and unloading process.[154] It is this exception to the exception that puts Himalaya clauses on a statutory footing.[155]

It should be noted that the contracting parties may exclude the application of the 1999 Act.[156] However, section 7(1) provides that the 1999 Act does not affect any other right or remedy available to a third party, thus the 1999 Act does not preclude the application of the other legal bases for the enforcement of Himalaya clauses. The test allowing third parties to enforce contractual clauses is provided for in section 1:

> (1) Subject to the provisions of this Act, a person who is not a party to a contract (a "third party") may in his own right enforce a term of the contract if—
>
> (a) the contract expressly provides that he may, or

[152] *The Mahkutai* (n 106) 8.
[153] Law Commission Report (n 12).
[154] Explanatory notes to the Contracts (Rights of Third Parties) Act 1999, para 26.
[155] Tetley, 'The Himalaya clause – Revisited' (n 121) 15.
[156] *Secure Capital SA v Credit Suisse AG* [2017] EWCA Civ 1486.

> (b) subject to subsection (2), the term purports to confer a benefit on him.
>
> (2) Subsection (1)(b) does not apply if on a proper construction of the contract it appears that the parties did not intend the term to be enforceable by the third party.
>
> (3) The third party must be expressly identified in the contract by name, as a member of a class or as answering a particular description but need not be in existence when the contract is entered into.

Section 1 of the 1999 Act allows a third party to enforce a contractual clause in two scenarios: (a) when there is an express contractual provision to this effect; or (b) when there is a contractual clause that purports to confer a benefit on the third party and there is nothing in the construction of the contract that indicates that the parties did not intent to allow the third party to enforce it. In both scenarios, the third party must be identified by name, class or description.[157] The identification of a third party during negotiations will not suffice.[158] However, the third party does not need to exist when the contract is made, thus the carrier's present and future subcontractors may qualify.[159]

The second scenario is a revolutionary development for third party protection in English law because it allows a third party to enforce a contractual clause that purports to confer a benefit on it, even if the contract does not give the third party a right to enforce. In other words, the conferral of a benefit on a third party by a contractual clause implies its right to enforce it.[160] Sections 1(1)(b) and 1(2) create a presumption that third parties are generally entitled to enforce a contractual clause if the clause confers a benefit on them and the contract expressly

[157] See *Laemthong InternationalLines Co Ltd v Artis & Ors* [2004] EWHC 2226 (Comm) [47] (it was construed that the term "agents" in the contract identified third party shipowners by class); *Starlight Shipping Co v Allianz Marine and Aviation Versicherungs AG & Ors* [2014] EWHC 3068 (Comm) [87], [88] (it was construed that the term "underwriters" in the contract identified the servants and agents of the underwriters by class, including their solicitor and loss adjuster); *The Prudential Assurance Co Ltd v Ayres & Ors* [2007] EWHC 775 (Ch), [2007] 3 All ER 946 [27] (it was construed that the term "any previous tenant" in the contract identified the underlessee of the premises by description); *Themis Avraamides & Anor v Colwill & Anor* [2006] EWCA Civ 1533 [17] (it was construed that the contract did not expressly identify any third party or class of third parties).

[158] Beatson, Burrows and Cartwright (n 11) 626.

[159] ibid.

[160] See Corcione (n 10) para 7.97 (Himalaya clauses and agency may be no longer necessary because the consent of the parties to confer a benefit on a third party by a contractual term may be adequate to allow the third party to enforce it).

identifies them by name, class or description.[161] This presumption is rebutted if, on a proper construction of the contract as a whole and the surrounding circumstances, the parties had not intended the third party to have the right to enforce the clause.[162]

Phrased in the shipping context: the second scenario allows a stevedore to enforce an exemption clause in the bill of lading between carrier and shipper, without an explicit provision to this effect, provided that (i) the exemption clause purports to protect the stevedore against liability, (ii) there is no contractual indication that the parties did not intent to entitle the stevedore to enforce, and (iii) the contract expressly identifies the stevedore by name, class or description.

(ii) The Canadian Test

In Canada, the test for the enforcement of Himalaya clauses was introduced by the Supreme Court in *London Drugs Ltd v Kuehne & Nagel International Ltd*.[163] The facts were that the appellant delivered to a warehouse company a transformer for storage which was damaged due to the negligence of the respondents, the company's employees.[164] Section 11 of the storage contract between the appellant and the employer of the respondents limited the liability of the employer to $40 per package unless the appellant paid the additional warehouse liability insurance.[165] The appellant, having full knowledge and understanding of this provision, chose not to obtain this insurance and instead arranged for its own coverage.[166] The issue was whether the employees, who were third parties to the storage contract, could rely on section 11 to exclude their liability in negligence.

The Court noted that the traditional exceptions to the privity rule, for example agency or trust, did not apply to the present case and, instead of artificially extending them beyond their accepted limits like in *The Eurymedon* and *ITO*, it was preferable to address the matter differently.[167] The Court ruled that the present case called for the relaxation of the privity rule in order to conform with the intentions of the contracting parties,[168] commercial reality,[169] as well as to prevent the appellant from circumventing the limitation clause to which it had

[161] Beatson, Burrows and Cartwright (n 11) 627. See also *Soprim Construction SARL v Republic of Djibouti* [2016] EWHC 3864 (Comm) [9].
[162] Beatson, Burrows and Cartwright (n 11) 627.
[163] *London Drugs Ltd v Kuehne & Nagel International Ltd* [1992] 3 SCR 299.
[164] ibid [2], [155].
[165] ibid.
[166] ibid [156].
[167] ibid [236].
[168] ibid [240], [245].
[169] ibid [212], [231], [240], [252], [268].

expressly agreed by suing the employees.[170] Thus, the Court adopted the following new test that allows employees to benefit from a limitation clause in a contract between their employer and its customer:

> 1. The limitation of liability clause must, either expressly or impliedly, extend its benefit to the employees (or employee) seeking to rely on it; and
>
> 2. the employees (or employee) seeking the benefit of the limitation of liability clause must have been acting in the course of their employment and must have been performing the very services provided for in the contract between their employer and the plaintiff (customer) when the loss occurred.[171]

Under this test, an employee may enforce an exemption or limitation clause in a contract between its employer and another party, if the exemption or limitation clause expressly or impliedly includes the employee and the employee was acting in the course of its employment and performing the employer's obligations under the contract when the loss occurred.

In applying this new test to the facts, the Court found that both requirements are met. Under the first requirement, although the contracting parties did not expressly include the word "employees" in the limitation clause,[172] there was nothing in the contractual language and all relevant circumstances precluding the employees from taking advantage of the limitation clause.[173] Consequently, the employees were implied third party beneficiaries of the limitation clause.[174]

The Supreme Court stated that this new test is similar to the agency test, as set out in *Scruttons*, because the first requirement in both tests is identical, the second and third requirements of the agency test are replaced by the identity of interest between employers and employees, and the fourth requirement of the agency test and the second requirement of the new test require the same reasoning with Himalaya clauses.[175] The Court also clarified that the new exception does not exclude the application of other exceptions to privity in case where the requirements of the new exception are not met.[176]

170 ibid [246], [250].
171 ibid [257].
172 ibid [265].
173 ibid [264], [266].
174 ibid [265].
175 ibid [259].
176 ibid [261].

Later in *Fraser River Pile & Dredge Ltd v Can-Dive Services Ltd*,[177] the Supreme Court named the new test set out in *London Drugs* as a "principled exception" to the privity rule which applies not only to the employee-employer relationship but to any contract that confers a benefit on a third party,[178] including a contract for the carriage of goods.

The decision in *London Drugs* has been applauded for noting the criticism of English judges in the development of the Himalaya clause, as well as the commercial reality that third party beneficiaries should not be "thwarted by legal niceties" from relying on a contractual clause.[179]

(iii) Similarities

Admittedly, both tests provide a simpler mechanism for enforcing Himalaya clauses in bills of lading since they exclude technicalities such as agency, ratification, and consideration.[180]

Both the English and Canadian tests are exceptions to the privity rule.[181] It should be noted however that the 1999 Act has been considered as an abolition of the privity rule,[182] whereas the principled exception has been considered as the result of an incremental change to the common law.[183] Although the English test is a statutory exception and the Canadian test is a common law exception, they both allow a person to enforce a provision in a contract to which it is not party, subject to certain requirements. It should be noted that both exceptions have altered the privity rule only to the extent that a contract may confer enforceable benefits on third parties, but they nevertheless left intact the part of the privity rule providing that a contract cannot impose obligations on third parties.[184]

The first requirement of the Canadian test is identical with section 1(1)(b) of the 1999 Act in that, where a contractual term expressly or impliedly confers a benefit on a third party, the third party may enforce such term. Emphasis should

[177] *Fraser River Pile & Dredge Ltd v Can-Dive Services Ltd* [1999] 3 SCR 108 [23], [31]-[39] (the respondent could rely on a waiver of subrogation clause in the contract between appellant and insurer, subject to the principled exception).

[178] ibid [31]–[32].

[179] Zeller and Moens (n 10) 262-263. However, see Neyers (n 148) 768-774 arguing that the "intentions of the parties" and "commercial reality" reasoning in *London Drugs* (n 163) and *Fraser* (n 177) is inadequate.

[180] Tetley, 'The Himalaya clause – Revisited' (n 121) 16, 19.

[181] Finch and Horsman (n 41) 415; Law Commission Report (n 12) para 2.8-2.18, 5.16 ("Our proposed statute carves out a general and wide-ranging exception to the third party rule"); *Halsbury's Laws of England* (5th edn, 2019) vol 22, para 143.

[182] *Alfred McAlpine Construction Ltd v Panatown Ltd* [2000] UKHL 43, [2001] 1 AC 518 (Lord Clyde).

[183] *Fraser* (n 177) [30] (not "a wholesale abdication of existing principles"); *London Drugs* (n 163) [254], [257].

[184] Beatson, Burrows and Cartwright (n 11) 633; Michael Bridge, 'The Contacts (Rights of Third Parties) Act 1999' (2001) 5 Edin L Rev 85, 85-86; Neyers (n 148) 765; *London Drugs* (n 163) [200].

be given on three identical points: First, the benefit must be conferred on a third party by a *contractual term* and not by the contract as a whole. Second, the conferral of a benefit on a third party by a contractual term equates its right to enforce it. Third, the phrase "expressly or impliedly" in the Canadian test has the same meaning with the phrase "purports to confer a benefit" under the English test. The word "purports" in the English test has been interpreted as signifying the intention of the contracting parties to confer a benefit on a third party either explicitly or implicitly.[185]

Furthermore, both tests have shifted the traditional presumption that third party beneficiaries cannot enforce contractual terms. The current presumption is that third party beneficiaries are entitled to enforce a contractual term unless the parties intended otherwise. Attention should be given to the negative wording of section 1(2) of the 1999 Act which provides that section 1(1)(b) does not apply if it is found, on a proper construction of the contract, that the parties "did not" intend the third party to enforce. This shows that if the contract is neutral on whether the parties intended to give a right of enforcement, section 1(1)(b) applies.[186] Similarly, in *London Drugs*, the employees were implied beneficiaries because there was nothing in the contract or the surrounding circumstances precluding the employees from relying on the limitation clause.[187]

(iv) Differences

The first major difference between the English and the Canadian test is that the 1999 Act distinguishes between the conferral of a benefit on the third party and the right of the third party to enforce the benefit,[188] while the principled exception does not make such distinction. In the Canadian test, the right of the third party to enforce a contractual term depends on the conferral of a benefit. However, section 1(1)(a) of the 1999 Act provides that a third party may enforce a contractual term where the contract expressly provides that it may, irrespective of whether the term confers a benefit on the third party or not, for example a jurisdiction clause, time limitation clause, and any other clause that does not necessarily confers a benefit on a third party. Of course, in contracts of carriage of goods by sea, where the 1999 Act only allows the enforcement of exemption or limitation clauses, this distinction is not so acute.

The second difference is that the English test does not require the third party to prove that it was acting in the course of its employment and performing

185 Bridge (n 184) 89.
186 *Nisshin* (n 124) [23].
187 *London Drugs* (n 163) [264], [266].
188 Bridge (n 184) 89.

the services provided in the contract of carriage when the loss or damage occurred, as required by the second requirement of the Canadian test. It is presumed that the 1999 Act applies only to the period specified in the bill of lading, *i.e.* usually after loading and before discharge, but this is not clear from the statutory provisions. It is further presumed that a shipper, who brings claims against a third party for damage of cargo, may argue that the third party cannot enforce a contractual clause under section 1 of the 1999 Act because it was acting outside the scope of the bill of lading when the damage occurred, and therefore this requirement will become relevant. But in such circumstances, it is likely that the burden of proof will fall on the shipper and not on the third party. In contrast, under the Canadian test, it is the third party that must prove that both requirements are met.

Another difference is that section 1(3) of the 1999 Act requires the third party to be expressly identified in the contract by name, class, or description, while the Canadian test does not. This means that the 1999 Act may give the right of enforcement to unnamed third parties but not to implied third parties.[189] Emphasis should be given to section 1(3) of the 1999 Act which requires the identification of a third party in the contract and not in the particular term on which the third party wishes to rely. Hence, if a stevedore wishes to enforce an exemption clause in a bill of lading, it must be expressly identified by name, class, or description anywhere in the bill of lading, and not necessarily in the exemption clause.

On the contrary, the Canadian test may give the right of enforcement to implied third parties. In this regard, it has been argued that the implication of an intent to benefit a third party may slide into fiction.[190] It is not easy to infer whether the failure of the contracting parties to include their employees in the contract was an oversight or a drafting glitch, or whether it shows that they did not intent to include them.[191] It should be reminded that Lord Denning in *Adler* ruled that carriers may stipulate an exemption clause for themselves and any third party engaged to perform the contract, either expressly or by necessary implication, provided that the shipper assented to the exemption, either expressly or by necessary implication.[192] Thus, where a contract does not explicitly protect a third party against liability, courts may imply the intent to protect the third party only if it is necessary to give business efficiency to the contract. Otherwise, courts would import to the contract a benefit that was not there nor intended to be there. Although it may be easier to imply that the parties clearly intended to protect their

[189] ibid 87-88. See *Themis* (n 157) [19] ("section 1(3), by use of the word 'express', simply does not allow a process of construction or implication").

[190] Fleming (n 22) 437.

[191] ibid.

[192] *Adler* (n 17) [11].

employees under the contract, as employees are often responsible for performing the contractual obligations of their employer to the plaintiff's knowledge,[193] it may be more problematic to prove that the parties intended to protect independent contractors.[194] As discussed above at Section II.A.(i), stevedores are independent contractors and not employees of the carrier, therefore the implication of an intent to benefit independent contractors is unlikely to be necessary for business efficiency. Consequently, the application of the Canadian test to contracts of carriage requires prudent reasoning.

Another difference is that the 1999 Act clarifies that a third party's right of enforcement may be subject to certain conditions. Section 1(4) of the 1999 Act provides that a third party may enforce a contractual term "subject to and in accordance with any other relevant terms of the contract". Thus, if a benefit is conferred subject to a qualification or condition, the third party must meet such qualification or condition in order to enforce the benefit.[195] For example, a third party seeking to bring claims against a contracting party is bound by the time limitation clause in the contract.[196] Another example is that, under section 8 of the 1999 Act, a third party seeking to enforce a contractual term is bound to do so in arbitration if any dispute arising out of or in connection with such term is subject to arbitration.[197] However, it may be difficult to distinguish between conditional benefits and obligations, especially where the benefit is conditional upon performance by the third party,[198] for example where a stevedore may benefit from an exemption clause subject to the condition that it performs the loading, or worse, where a stevedore may benefit from an exemption clause subject to the condition that it *successfully* performs the loading. In such cases, the conferral of a conditional benefit may be indistinguishable from the imposition of an obligation. Having said that, it is interesting to consider whether a third party could set aside a condition, for example an arbitration agreement, on the basis that it is unreasonable or unconscionable.[199] It has been argued that the third party should be bound by a condition in the limited sense that the promisor (the shipper) may use the condition as a defence to a claim brought by the third party to enforce the

193 Fleming (n 22) 430, 433.

194 Beatson, Burrows and Cartwright (n 11) 651, note 27.

195 ibid 633.

196 Bridge (n 184) 93.

197 *Nisshin* (n 124) [34], [39]; Explanatory Notes to the Contracts (Rights of Third Parties) Act 1999, para 33-35; However, see *The Mahkutai* (n 106) 13-14, where the Privy Council did not allow a third party to enforce an exclusive jurisdiction clause in the bill of lading because jurisdiction clauses, contrary to exemption clauses, embody a mutual agreement between the contracting parties with mutual rights and obligations, thus they are not included in a contract for the sole benefit of the carrier and its subcontractors.

198 Beatson, Burrows and Cartwright (n 11) 633.

199 See for example *Uber Technologies Inc v Heller* [2020] SCJ No 16.

contract.[200] The decision in *London Drugs* does not regulate this issue but it is nevertheless presumed from the reasoning in *Fraser* that a third party may enforce a contractual clause subject to any express qualifying language or limiting conditions in the contract.[201] But this remains to be examined in the common law.

Finally, the decisions in *London Drugs* and *Fraser* suggest that the Canadian test can only be used as a shield and not as a sword.[202] This means that while third parties can rely on exemption or limitation clauses to defend themselves against the claims of a contracting party, they cannot sue the contracting party on the contract.[203] In contrast, the 1999 Act seems to allow third parties to bring an action against contracting parties.[204] However, since the 1999 Act only allows the enforcement of exemption or limitation clauses in contracts of carriage of goods, it is likely that third parties will only use the 1999 Act as a shield.

(v) Discussion

Both tests put subcontractors in a more favourable position. The effect of the tests is substantially similar. The Canadian test is deemed to be a wider exception to the third-party rule than the English test as it does not require the third party to be expressly identified by name, class or description.[205] However, the 1999 Act provides a more complete and comprehensive test for third party rights in that it puts Himalaya clauses on statutory footing and regulates the contracting parties' rights as well.

The disadvantage of both tests, however, is that they do not specifically address the contemporary needs of the shipping industry. Particularly, it may be problematic that neither test regulates multimodal contracts of carriage. In multimodal transport, there is a greater number of third parties involved in the chain of carriage, including maritime parties (ocean carriers, charterers, stevedores) and non-maritime parties (rail/road/air carriers, terminal operators, freight forwarders) and their subcontractors. Can all these third parties benefit from a limitation clause in the principal contract of carriage? The English test allows the enforcement of a limitation clause by third parties engaged in the

[200] Beatson, Burrows and Cartwright (n 11) 633.

[201] *Fraser* (n 177) [42].

[202] *London Drugs* (n 163) [260]; *Fraser* (n 177) [37]; Neyers (n 148) 765.

[203] ibid. However, see *Brown v Belleville (City)* [2013] OJ No 1071 [110]-[111] where the successors of the original covenantee had a right to enforce the agreement against the original covenantor.

[204] Explanatory notes to the Contracts (Rights of Third Parties) Act 1999, para 15; Law Commission Report (n 12) para 11.14, 11.19; Treitel (n 15) 104-105; Beatson, Burrows and Cartwright (n 11) 634 (in actions brought by third parties, the promisor is entitled to raise any defence or set-off under the contract, eg mistake, misrepresentation, breach, frustration).

[205] Beatson, Burrows and Cartwright (n 11) 651.

loading and unloading process,[206] and the Canadian test allows the enforcement of a limitation clause by third parties performing the very services provided for in the contract.[207] This is not easily determined in shipping practice. In fact, it is extremely difficult to accurately define the activities and obligations of carriers in the modern context of carriage because modern carriers assume responsibility for a greater period of carriage, including maritime and inland operations.[208] Suppose that a limitation clause in a bill of lading limits the liability "of all servants, agents and independent contractors of the carrier (including their servants, agents and independent contractors) for any loss or damage of cargo that occurs in the period between loading and discharge" without defining these terms; it is unclear whether a third party may enforce the limitation clause if the loss or damage occurred while loading or during storage or while unloading from truck or rail. This is a matter of interpretation and previous judicial decisions may provide some guidance.[209] Nevertheless, neither test draws a line between maritime and non-maritime operations. Thus, any third party engaged in a multimodal contract of carriage can presumably enforce a limitation clause, even if it is remotely connected (if connected at all) with the principal contract. This may expose the shipper to a high risk of not recovering its losses from negligent third parties in any mode of carriage.

Of course, the English and Canadian tests have not yet been widely applied in the carriage of goods context. It remains to be seen in the common law whether the tests constitute a satisfactory legal basis for the enforcement of Himalaya clauses or whether the shipping industry requires a more targeted test.

III. Conclusion

The development of Himalaya clauses in England and Canada has been very similar. Since Canadian maritime law applies, if not exclusively, the common law of England,[210] the Supreme Court of Canada has consistently affirmed and applied English decisions on the matter.

In both jurisdictions, the enforcement of Himalaya clauses has been incremental due to the factually-limited opportunities given to courts to determine

[206] Explanatory notes to the Contracts (Rights of Third Parties) Act 1999 para 26.

[207] *London Drugs* (n 163) [257].

[208] Corcione (n 10) para 7.94.

[209] See *Raymond Burke* (n 118) (the dock operators negligently damaged the cargo in the container park before it passed the ship's rail, and the court interpreted whether the exemption clause in the bill of lading covered that period); The Rigoletto (n 118) (the stevedores received the cargo some days before the scheduled loading and it was stolen during storage, and the court interpreted whether the exemption clause in the bill of lading covered that period); *The New York Star* (n 98)*; ITO* (n 46).

[210] William Tetley, 'A Definition of Canadian Maritime Law' (1996) 30 UBC Law Rev 137, 146.

the issue in the context of carriage of goods by sea (almost one case per decade). Since the strict application of the privity rule in *Scruttons* in 1961, it has taken the English Parliament 38 years to regulate third party protection by statute until the enactment of the 1999 Act. Similarly, in Canada, since the adoption of *Scruttons* in *Canadian General Electric* in 1971, it has taken the Supreme Court 28 years until *Fraser* in 1999, which extended the application of the principled exception to *inter alia* contracts of carriage.

It appears from the analysis that English and Canadian courts have struggled in finding a suitable and widely accepted legal basis for the enforcement of Himalaya clauses, but they had been generally willing to reach a commercially sensible solution. In fact, it was commented that the reasoning of the Supreme Court in *London Drugs*, as well as the close analysis of precedents, reminds of Lord Denning's "American" style that pays more attention to the actual decision than the explanation.[211] Indeed, it is evident from both English and Canadian decisions that courts were more interested in "getting" to the enforcement of Himalaya clauses rather than in providing prudent reasoning for it.

This article underlines the perplexity of each legal basis for the enforcement of Himalaya clauses. Based on the foregoing analysis, none of the legal bases applies to the shipping context in an adequate and generally accepted manner. On the contrary, each legal basis upon which Himalaya clauses have been enforced highlights the tension between commercial practice and existing legal principles. The agency basis, for example, does not easily apply to the commercial relationship between carrier and third parties, and the unilateral contract basis give rise to issues relating to the provision of consideration. Moreover, the exemption clause basis, together with the broad interpretation adopted by English and Canadian courts, entails the risk of implying words to the contract that are simply not there. In response to these issues, England and Canada eventually proceeded with the establishment of new generic tests that allow third party beneficiaries to enforce contractual provisions. Currently, in both jurisdictions, the enforcement of Himalaya clauses does not depend on technicalities such as agency, ratification, offer and consideration, but on contract drafting and the parties' intentions. Nevertheless, the 1999 Act in England and the principled exception in Canada present some important differences which raise concerns as to the extent of the right of third parties to enforce contractual provisions. Thus, the effectiveness of the legal bases for the enforcement of Himalaya clauses is not uncontested.

Whether the 1999 Act in England or the principled exception in Canada is the final home of Himalaya clauses remains to be seen in the common law. As already mentioned, disputes over the enforcement of Himalaya clauses are very

[211] Fleming (n 22) 438-439.

sporadic in litigation. Having in mind that both tests are dated in the 1990s and do not sufficiently address the current legal issues in the context of multimodal carriage of goods, it is presumed that they are not the ultimate legal basis for the enforcement of Himalaya clauses. Let us hope that it will not take another thirty years to find out.

Cambridge Law Review (2022) Volume VII, Issue 1, 121–154

Collective Bargaining Trends in Nigeria—Living up to the International Labour Organisation (ILO) Standards?

SAMSON FAITHFUL OBIORA*

ABSTRACT

Collective bargaining is one of the most observed and least regulated phenomena in labour and industrial relations. Labour laws touching on industrial relations and collective bargaining in Nigeria are devoid of codification and found scattered across our statute books. A call for a focus on collective bargaining is no doubt apposite and topical reflecting the worldwide thrust towards fundamental freedoms and trade unionisation. The ubiquity of collective bargaining practices has made international organisations, like the ILO to become "negotiation infatuated" by giving standard prescriptions and insisting that "voluntarism" is the key framework for the viability of collective bargaining. Nonetheless, after five decades of Nigeria's membership with the ILO and ratification of its human rights instrument, how much have its industrial relations and collective labour policies improved? Nigerian workers continue to wallow in the shadow of their organisational rights, and indeed the spatial culture of interventionism and compulsion in Nigeria's regulatory landscape. This study negates the perspectives that prioritize administrative intrusion at the expense of commitment to voluntarism. The study engages in a comparative critique of Nigeria's collective bargaining framework vis-a-vis the benchmark labour standards of the ILO. Additionally, the study considered collective bargaining in a comparative approach with the United Kingdom and South Africa jurisdictions focusing on the extent of legislative recognition of the duty to bargain and the enforcement of the collective agreement as a finished product of the bargaining process. Part of the findings was that whereas there is neither a statutory obligation to bargain nor are collective agreements readily enforceable in Nigeria, in other jurisdictions the

* BL Candidate at the Nigerian Law School, Lagos, Nigeria. LLB (University of Lagos). I am grateful to the anonymous reviewers for their comments on earlier drafts. Any errors that remain are my own.

right bargain is accorded statutory flavour, and collective agreements in so far as the parties to it intend that the agreement should bind them, it is enforceable. Beyond this, the study under the themes of "legal frameworks" and "governing principles" of the right to organise, reveals the inherent challenges of collective bargaining in Nigeria. This study in panoramically reflecting on the standard prescriptions of the ILO and key collective bargaining indicators of other jurisdictions, suggests policy reform as a panacea to bridge the lacuna and pace up the lag behind international labour standards.

Keywords: collective bargaining; collective agreement; Nigeria; International Labour Organisations (ILO); United Kingdom; South Africa

I. INTRODUCTION

Collective bargaining is a process of negotiation and conclusion of collective agreements on terms and conditions of employment between employers and employees.[1] It is an important mechanism for attaining a cordial relationship between workers and their employers because it provides an effective forum for the settlement of employment issues.[2] In broad terms, Davey has defined collective bargaining as a constitutional relationship between an employer entity (government or private) and labour organisation (union or association) representing exclusively, a defined group or employees of said employer (appropriate bargaining unit) concerned with the negotiation, administration, interpretation and enforcement of written agreement covering joint understanding about wages or salaries, rate of pay, hours of work and other conditions of employment.[3]

In terms of the International Labour Organisations Law (ILO Law),[4] collective bargaining is explained as extending to all negotiations which take place between an employer, a group of employers or one or more employers' organisations, on the one hand, and one or more workers' organisations, on the other hand, for determining working conditions and terms of employment, and regulating relations between employers and workers, and regulating relations between employers or their organisations and a workers' organisation or workers' organisations.[5] Collective bargaining thus involves a situation where representatives of organised employees meet with the employer or its represen-

1 OVC Okene, 'The Challenges of Collective Bargaining in Nigeria: Trade Unionism at the Cross-Roads' (2010) 4 Labour Law Review 61.

2 ibid.

3 Harold W Davey, *Contemporary Collective Bargaining* (3rd edn, Prentice Hall Inc 1972) 64.

4 The expression "ILO Law" is used here as a generic term for its Convention.

5 Collective Bargaining Convention (No 154) 1981, art 2.

tatives in an atmosphere of mutual cooperation and respect, to deliberate and reach agreement on issues affecting both parties.[6]

The International Labour Organisation (ILO), as the pre-eminent body on international labour standards, has by its Conventions and Recommendations provided the legal framework to guide Member States to enact domestic laws and provide mechanisms to facilitate the practice of collective bargaining.[7] Nigeria is a member of the ILO[8] and she has ratified both the ILO Freedom of Association and Protection of the Right to Organise Convention (No 87) 1948, and the Right to Organise and Collective Bargaining Convention (No 98) 1949.[9]

Profound as the above may seem, in derogation of these core labour standards, Nigerian workers continue to lack these basic rights. The standard principles that underlay the practice of collective bargaining have been applied differently. In Nigeria, neither the Constitution[10] nor the Labour Act[11] is characterised with the recognition of a statutory duty to bargain.[12] The legal draftsmen have opted for a paradigm which allows the social partners through the exercise of power, to resolve their own arrangements. The power play is given legal impetus by the provisions on condition of employment[13] vis-à-vis the protected right to freedom of association[14] and the recognition of trade unions.[15] Likewise, it has been expressed that no Nigerian legislation clearly defined the term "collective bargaining".[16] The Trade Disputes Act[17] and the National Industrial Court Act[18] merely defined collective agreement. It needs be added as

6 Robinson Olulu and Sylvester Udeorah, 'The Principle of Collective Bargaining in Nigeria and the International Labour Organisation (ILO) Standards' (2018) 3 International Journal of Research and Innovation in Social Science 63.

7 Okene, 'The Challenges of Collective Bargaining in Nigeria' (n 1) 62.

8 'Country Profile' (*International Labour Organisations Normlex*, 20 October 2021) <https://www.ilo.org/dyn/normlex/en/f?p=NORMLEXPUB:11003:0::NO:::#M> accessed 20 October 2021.

9 Both conventions were ratified on 17 October 1960. See 'Ratifications for Nigeria' (*International Labour Organisations Normlex*, 20 October 2021) <https://www.ilo.org/dyn/normlex/en/f?p=1000:11200:0::NO:11200:P11200_COUNTRY_ID:103259> accessed October 2021.

10 Constitution of the Federal Republic of Nigeria 1999.

11 Cap L1 LFN 2004.

12 In South Africa, which is close to Nigeria in more ways than one, section 23(5) of its National Constitution (No. 108 of 1996) confers the right to collective bargaining. It provides expressly that 'Every trade union, employers' organisation and employer has the right to engage in collective bargaining. National legislation may be enacted to regulate collective bargaining. To the extent that the legislation may limit a right in this Chapter, the limitation must comply with section 36 (1)'.

13 Labour Act, s 9 (6).

14 Constitution of the Federal Republic of Nigeria 1999, s 40.

15 Section 25 of the Trade Unions Act, Cap T14, LFN 2004. See *Mix and Bake Flour Mill Industries Ltd v National Union of Food, Beverage and Tobacco Employees* [1978-2006] DJNIC 277.

16 Richard Idubor, *Employment and Trade Disputes Law in Nigeria* (Sylva Publishers Ltd 1999) 40.

17 CapT8 LFN, 2004.

18 National Industrial Court Act 2006.

the correct position, that although not elaborate, the Nigerian Labour Act defines collective bargaining as "the process of arriving at or attempting to arrive at, a collective agreement".[19] Against this background, the objective of this paper is to set out the ILO's principles of collective bargaining as they emerge from the various legislative frameworks adopted by the Organisation and the comments made by its supervisory bodies — notably the Committee of Experts on the Application of Conventions and Recommendations (CEACR) and the Committee on Freedom of Association (CFA) vis-a-vis its empirical application in Nigeria, and its implications for Nigerian workers.

Consequently, Part I defines the "collective bargaining" concept and introduces the central theme of the paper. Part II sets out the legal frameworks for collective bargaining in the context of ILO standards. This part via a comparative analysis examines the issues relating to parties to collective bargaining; the recognition of workers' organisations; employees and subject matters covered by collective bargaining; and the choice of bargaining level. Part III examines the governing principles of ILO standards in terms of the principles of free and voluntary negotiation, good faith and the enforcement of collective bargaining agreements. Part IV analyses collective bargaining in a comparative approach with other jurisdictions. Part V provides conclusion to the study and recommends amongst others, the need for a legal reform in Nigeria in a bid to pace up the lag behind international standards.

II. The Right to Collective Bargaining And The ILO Standard Legal Frameworks

The ILO is the supreme authority on international labour standards. The ILO provides the major human rights instrument that guarantees and advances organisational rights[20] and has carried out an enormous amount of standard-setting work during the 80 years of its existence as it has sought to promote social justice, and one of its chief tasks has been to advance collective bargaining throughout the world.[21] This task was already laid down in the Declaration of Philadelphia, 1944, part of the ILO Constitution, which stated "the solemn obligation of the International Labour Organisation to further among the nations

[19] Labour Act, s 91.

[20] Okene, 'The Challenges of Collective Bargaining in Nigeria' (n 1) 64.

[21] Nicholas Valticos, 'The ILO: A Retrospective and Future View' (1996) 135 International Labour Review 473.

of the world programmes which will achieve...the effective recognition of the right of collective bargaining".[22]

Three major international instruments have been adopted by members of the ILO with the aim of promoting collective bargaining amongst member states. These are the following: (a) Freedom of Association and Protection of the Right to Organise Convention (No 87)[23] (b) The Right to Organise and Collective Bargaining Convention (No 98)[24] and (c) Collective Bargaining Convention (No 154).[25]

In 1948 the ILO adopted Convention No 87 on Freedom of Association and Protection of the Right to Organise. This Convention established the right of all workers to form and join organisations of their own choosing, and set out guarantees for workers' organisations to function independently of government control.[26] These organisations shall also have the right to establish and join federations and confederations and affiliate with international organisations of workers.[27] There are also guarantees ensuring the right of workers' organisations to function freely.[28] Furthermore, Member States are under an obligation to take all necessary and appropriate measures to ensure that workers may exercise freely the right to organise; and the law of the land shall not be such to impair nor shall it be applied to impair the guarantees provided in the Convention.[29] The Convention further clarifies that national legislation shall determine the extent to which this Convention shall apply to the armed forces and the police.[30] Convention No 87 has been described as "the most comprehensive international instrument in this area of human rights and has become a pivotal reference point within the broad area of trade union law and practice".[31]

Furthermore, ILO Convention No 98 (1949) on the Right to Organise and Collective Bargaining goes on to protect, workers against acts of anti-union discrimination in respect of their employment.[32] The workers' organisations are also protected against interference by other organisations and by employers in

[22] ILO Constitution of the International Labour Organisation and Standing Orders of the International Labour Conference 1998.

[23] Freedom of Association and Protection of the Right to Organise Convention (adopted 9 July 1948, entered into force 4 July 1950) C087.

[24] The Right to Organise and Collective Bargaining Convention (adopted 1 July 1949, entered into force 18 July 1951) C098.

[25] Collective Bargaining Convention (adopted 3 June 1981, entered into force 11 August 1983) C154.

[26] Convention No 87 (n 23), art 3.

[27] ibid art 5.

[28] ibid art 2.

[29] ibid arts 8 and 11.

[30] ibid art 9.

[31] Von G Potosbsky, 'Freedom of Association: The Impact of convention 87 and ILO Action' (1998) 137 International Labour Review 1.

[32] Convention No 98 (n 24), art 1.

their establishment, function and administration.[33] Additionally, the Convention provides for the obligation to establish machinery appropriate to national conditions, where necessary to ensure respect for the right to organise and encourage the full development and utilisation of the machinery for collective bargaining.[34] Similar to Convention No 87, the application of Convention No 98 to the armed forces and the police depends on national legislation.[35] Convention No 98 also does not deal with the position of public servants engaged in the administration of the State.[36] These two Conventions were followed in 1981 by the Collective Bargaining Convention No. 154 which also promotes free and voluntary collective bargaining.[37]

More recently, in June 1998, the ILO took another step forward by adopting the Declaration on Fundamental Principles and Rights at Work and its Follow-up.[38] This states that

> All Members, even if they have not ratified the [fundamental] Conventions, have an obligation, arising from the very fact of membership in the Organisation, to respect, to promote and to realize, in good faith and in accordance with the Constitution, the principles concerning the fundamental rights which are the subject of those [fundamental] conventions.[39]

The fundamental rights referred to in the Declaration include freedom of association and the effective recognition of the right to collective bargaining.[40]

As the world-acknowledged specialist agency on labour matters, the ILO has since its inception been at the forefront of the crusade to protect workers.[41] The ILO realised that workers would remain powerless so long as they stood as

33 ibid art 2.
34 ibid arts 3 and 4.
35 ibid art 5.
36 ibid art 6.
37 In addition to these Conventions, there are numerous Conventions and Recommendations which promotes collective bargaining between workers and their employers. These include Workers' Representative Convention (adopted 23 June 1971, entered into force 30 June 1971) C135 and Labour Relations (Public Service) Conventions (adopted 27 June 1978, entered into force 25 February 1981) C 151. Others include: Collective Agreement Recommendation (adopted 29 June 1951) R091; Voluntary Conciliation and Arbitration Recommendation (adopted 29 June 1951) R092; Collective Bargaining Recommendation (adopted 19 June 1981) R163.
38 H Kellerson, 'The ILO Declaration of 1998 on Fundamental Principles and Rights: A Challenge for the Future' (1998) 137 Internationalisation Labour Review 223.
39 International Labour Organisations, *Declaration on Fundamental Principles and Rights at Work and its Follow-up* (1st edn, Geneva 1998).
40 Bernard Gernigon and others, 'ILO Principles Concerning Collective Bargaining' (2000) 39 International Labour Law Review 34.
41 Okene, 'The Challenges of Collective Bargaining in Nigeria' (n 1) 68.

individuals in the face of heavily organised capital.[42] In this wise, Fox has opined that "The weakness of the individual worker makes the individual agreement for the sale of his labour power 'asymmetric', an exchange which cannot be gauged by reference to the so-called contract of employment".[43] This perhaps explains why the ILO is mostly concerned with the facilitation of individual workers to group together and found a force strong enough to bargain on equal terms with the employer and where necessary undertake industrial action to realise their demands.[44] Fully cognizant of the ILO's action affirming collective bargaining as a fundamental human right, the World Trade Organisation (WTO) in 1996 issued the following Ministerial declaration on core labour rights: "We renew our commitment to the observance of internationally recognised core labour rights. The ILO is the competent body to set and deal with these standards, and we affirm our support for its work in promoting them".[45] Without doubt, the ILO plays a crucial role in guaranteeing workers' rights and establishing a social framework that can ensure social justice throughout the world.[46]

A. Parties to collective bargaining and workers' organisation recognition

As a matter of labour practice, what makes the bargaining "collective" is the presence of a trade union(s) that represents the interests of employees as a collective entity.[47] The other party to collective bargaining is usually an employer. It could be a number of employers or an employer's organisation. At times, the government or a Government/State agency/institution could be the employer party as is the case in public service.[48] By the ILO standards, collective bargaining involves a bipartite relationship (between two parties). It does not extend to cover tripartite relations where the government is also a party. This is because the ILO

[42] ibid.

[43] Alan Fox, *Beyond Contract: Work, Power and Trust Relations* (Faber and Faber 1974) 191.

[44] As Morris noted, 'In the field of freedom of association the ILO has shown itself well able to appreciate the complexities of collective bargaining as demonstrated, in particular, by the principles it has developed in relation to industrial action'. See Gillian S Morris, 'Freedom of Association and the Interest of the State' in Keith D Ewing and others (eds), *Human Rights and Labour Law: Essays for Paul O'Higgins* (Mansell 1994) 51.

[45] Roy J Adams, *The Human right to Bargain Collectively: A Review of Documents supporting the International Consensus* (McMaster University1998).

[46] Isabelle Boivin and Alberto Odero, 'The committee of Experts on the Application of Conventions and Recommendations: Progress Achieved in National Labour Legislation' (2006) 145 International Labour Review 207.

[47] M-S Vettori, 'Alternative Means to Regulate the Employment Relationship in the Changing World of Work' (DPhil thesis, University of Pretoria 2005).

[48] Beverly M Musili, 'Challenges in Implementing and Enforcing Collective Bargaining Agreement' (2018) The Kenya Institute for Public Policy Research and Analysis Discussion Paper 208/2018, 9 <http://repository.kippra.or.ke/handle/123456789/2190> accessed 25 October 2021.

Convention on collective bargaining strives to create a balance between government intervention to encourage collective bargaining (by establishing an enabling framework) and the freedom of the parties to conduct autonomous negotiations.[49]

Specifically, the parties to collective bargaining are; one or more employers; or one or more employers' organisations on the one hand; and one or more workers' organisations on the other hand. Proceeding from a similar legislative approach, the Nigerian labour law reiterates this traditional bifurcated relationship in terms of its definition of collective agreement under the Trade Disputes Act and the National Industrial Court Act. For clarity, section 48 of the Trade Disputes Act defines "collective agreement" as

> any agreement in writing for the settlement of disputes and relating to the terms of employment and physical conditions of work concluded between an employer, a group of employers or organisations representing workers, or the duly appointed representative of any body of workers, on the one hand; and one or more of trade unions or organisations representing workers, or the duly appointed representatives of any body of workers, on the other hand.

In a similar vein, section 54 of the National Industrial Court Act defines "collective agreement" as:

> Any agreement in writing regarding working conditions and terms of employment concluded between
> a) an organisation of employers or an organisation representing employers (or an association of such organisation), of the one part, and
> b) an organisation of employees or an organisation representing employees (or an association of such organisation), of the other part.

It follows therefore that collective bargaining should be done between employers' organisations and workers' organisation.[50] In the absence of workers'

[49] Olulu and Udeorah (n 5) 64.

[50] The expression 'organisation' is used here as a generic term for federation of employers' organisations, individual employers' organisations, federation of trade unions and individual trade unions.

organisation, negotiations may be done with the workers' representatives. Nevertheless, where this is done "appropriate measures should be taken to ensure that the existence of these representatives is not used to undermine the position of the workers' organisations concerned".[51] The CFA maintained in one case that, 'direct negotiation between the organisation and its employees, by-passing representative organisations where these exist, might in certain cases be detrimental to the principle that negotiation between employers and organisations of workers should be encouraged and promoted.[52] It was also emphasised by the committee that "direct settlements signed between an employer and a group of non-unionised workers, even when a union exists in the undertaking does not promote collective bargaining as set out in Article 4 of Convention (No. 98)".[53]

At this juncture, an issue which needs to be examined is whether the right of the parties to negotiate is automatic upon the formation of workers' organisation or is subject to a certain level of representativeness. Strictly speaking, the requirement to be a registered organisation is the only condition laid down under the law. This undoubtedly seems to be a truism in the light of section 2 of the Trade Unions Act which prohibits unregistered trade union from functioning. Mere registration is not sufficient to entitle an organisation or a trade union to negotiate collective agreement within the meaning of the Act. The exercise of this privilege appears to be dependent on their recognition by the respective employer(s). It is trite that trade union recognition is germane to the very existence of workers' organisations. Freedom of association would be hollow and of no relevance to workers if employers were entitled to refuse to recognize their organisation for purposes of collective bargaining.[54] Thus union recognition is a sine qua non to collective bargaining.[55] Indeed the CFA has ruled that recognition by an employer of the main unions represented in his undertaking, or the most representative of these unions, is the very basis for any procedure for collective bargaining.[56] Where there is no union organisation in an industry, the position of the CFA is that the representatives of the unorganised workers duly elected and authorised by the workers will conduct bargaining on their behalf.[57] Under Nigerian labour law, as in the labour laws of other jurisdictions, the most

[51] This standard is set out in Paragraph 2 of Recommendation No 91 and is confirmed in Article 5 of Convention No 135. See also Convention No 154, art 3(2).

[52] International Labour Organisations, *Freedom of Association: Digest of Decisions and Principles of the Freedom of Association Committee of the Governing Body of ILO* (4th edn, Geneva 1996) para 785.

[53] ibid para 790.

[54] Joseph E Abugu, 'Democratic Trends in Industrial Relations: Progress and Drawbacks' (2012) 18 The Nigerian Journal of Contemporary Law 131.

[55] ibid.

[56] International Labour Organisations, *Freedom of Association: Digest of Decisions and Principles of the Freedom of Association Committee of the Governing Body of ILO* (5th edn, Geneva 2006) para 953.

[57] ibid, paras 785 and 786.

important step in the collective bargaining procedure is for the employer or the employees' association to recognize the trade union as a bargaining agent for the employees within the bargaining unit, in relation to terms and conditions of employment.[58] This is a matter of statutory obligation for employers, provided that a trade union has more than one of its members in the employment of an employer.[59]

Far reaching as the above may seem, such recognition per se does not confer an automatic right to bargain with individual unions. For the purpose of collective bargaining, all registered trade unions shall constitute an electoral college to elect members who will represent them in negotiations with the employer in collective bargaining.[60] In this wise, it may be argued that this obligation to negotiate collective agreements is reserved for the most representative organisations.[61] This requirement for an "electoral college" raises a number of drawbacks. The Trade Unions Act does not prescribe the modalities for constituting an electoral college. Perhaps it was thought that as democratic institutions, the unions should be able to work this out amongst themselves.[62] Such lapses do provide avenue for unfair employer interventions in the constitution of electoral colleges for collective bargaining. Put more specifically, this lacuna will have the tendency to encourage favouritism as employers will try to influence the criteria for the assessment of representatives, who would be disposed to management during negotiations.[63] The poser here becomes how exactly should the issue of representation be determined? In this respect, it should be recalled, depending on the individual system of collective bargaining, that trade union organisations which participate in collective bargaining may represent only their own members or all the workers in the negotiating unit concerned.[64] In this latter case, where a trade union (or, as appropriate, trade unions) represents the majority of the workers, or a high percentage established by law which does not imply such a majority, in many countries it enjoys the right to be the exclusive bargaining agent on behalf of all the workers in the bargaining unit.[65]

58 OVC Okene, 'The Internationalisation of Nigeria Labour Law: Recent Development in Freedom of Association' (2008) 7 University of Botswana Law Journal 93.

59 Trade Unions Act, s 24. See *Mix and Bake Flour Mill Industries Ltd. v National Union of Food, Beverage and Tobacco Employees (NUFBTE)* [1978-2006] DJNIC 277.

60 Trade Unions Act, s 24(1).

61 Abugu (n 54) 146.

62 ibid.

63 Okene, 'The Internationalisation of Nigeria Labour Law' (n 58) 103.

64 Gernigon and others (n 40) 34.

65 ibid 38.

Commenting on the issue of representativeness which the Act fails to prescribe, Abugu[66] and Okene[67] have opined that a better prescription would have been a "majoritarian" or "sufficiently representative" approach whereby the trade union with a majority of workers in the workplace or one which is sufficiently representative of the workers will be recognized to bargain on behalf of the workers. A criteria can be laid out for determining when a union is 'sufficiently representative' of workers, taking into account such factors as the size of the union, experience and contributions amongst other.[68] The principle of representativity ensures that employers do not find themselves in a position where they' are expected to include in negotiations every single trade union which has members, no matter how insignificant the membership.[69] Only those trade unions which could, to a large extent, influence relationship between employer and the body of employees within an agreed bargaining unit are to be allowed at the negotiation table.[70] All benefits accruing from the negotiations with management are enjoyed by all workers in the unit.[71] This is an accepted international law practice and is endorsed by the ILO Freedom of Association Committee. The Committee has in fact opined that the determination of such representation should be based on "objective and pre-established criteria" to avoid opportunity for partiality or abuse.[72] A legislative overhaul is therefore needed in Nigeria to provide an "objective and pre-established criteria" for determining representativity, and until such reform is made, suffice it to say that the issue of workers' organisations recognition is a huge challenge to collective bargaining practice in Nigeria and does not meet with the requirements of international practice.

B. Employees covered by collective bargaining

Generally, the Right to Organise and Collective Bargaining Convention (No 98) provides that workers shall enjoy adequate protection against acts of anti-union discrimination in respect of their employment.[73] It provides that "the extent to which the guarantees provided for in this Convention shall apply to the armed forces and the police shall be determined by national laws or regulations",[74] and also states that "this Convention does not deal with the position of public servants engaged in the administration of the State, nor shall it be construed as prejudicing

66 Abugu (n 54) 146.
67 Okene, 'The Internationalisation of Nigeria Labour Law' (n 58) 104.
68 Abugu (n 54) 146.
69 Okene, 'The Challenges of Collective Bargaining in Nigeria' (n 1) 82.
70 ibid 83.
71 ibid.
72 ibid para 962.
73 Convention No 98 (n 24) art 1.
74 ibid art 5.

their rights or status in any way".[75] Under this Convention, only the armed forces, the police and the above category of public servants may therefore be excluded from the right to collective bargaining. With regard to this type of public servants, the Committee of Experts has stated the following:

> The Committee could not allow the exclusion from the terms of the Convention of large categories of workers employed by the State merely on the grounds that they are formally placed on the same footing as public officials engaged in the administration of the State. The distinction must therefore be drawn between, on the one hand, public servants who by their functions are directly employed in the administration of the State (for example, in some countries, civil servants employed in government ministries and other comparable bodies, as well as ancillary staff) who may be excluded from the scope of the Convention and, on the other hand, all other persons employed by the government, by public enterprises or by autonomous public institutions, who should benefit from the guarantees provided for in the Convention.[76]

Proceeding from a similar legislative outlook, in Nigeria, the Trade Unions Act 1973 excluded from membership of trade unions staff recognized as "projection of management"[77] and certain class of public officers. And for the purpose of determining projection of management, section 3(4) of the Act provides that a person whose status, authority, powers, duties and accountability, as reflected in the conditions of service, are such as normally inhere in a person exercising executive authority, may be recognised as projection of management. Thus, staffs such as permanent secretaries, heads and secretaries of commissions, boards, companies and corporations — and, indeed, all staff in public and private establishment whose duties include policy and decision making, cannot be members of trade unions.[78] Within this category also fall all full professors who by their status have the right to become automatic members of Senate which alone takes decision on award of academic degrees and student discipline.[79]

[75] ibid art 6.

[76] International Labour Organisation, *General Survey Report on the Freedom of Association and Protection of the Right to Organise Convention 1948 (No 87) and the Right to Organise and Collective Bargaining Convention 1949 (No 98)* (Series B no 4, Geneva 1994) para 200. The Committee on Freedom of Association has made similar statements in the same vein. See ILO, Digest of Decisions and Principles of Freedom of Association Committee 1996 (n 52) paras 793-795 and 798.

[77] Trade Unions Act, s 3(3).

[78] Akintunde Emiola, *Nigerian Labour Law* (Emiola Publishers Limited 2008) 413.

[79] *Akintemise Ors v Onwumechili* [1985] ANLR 85.

Going forward, section 11(2) of the Trade Unions Act 1973 provides that "it shall not be lawful" for persons in the police, prison, and armed forces as well as those in the customs preventive services "to combine, organise themselves, or to be members of any trade union". So also are employees in the Nigerian Security Printing and Minting Company, staff of the Central Bank and the Nigerian External Telecommunications Limited, and those in any other services of the federal or state government "authorized to bear arms".[80] Additionally, the minister of labour is also empowered to specify "other establishments from time to time" whose staff may not belong to trade unions.[81] The Act preserves the right of such employees to take part in the setting up of joint consultative committees in the establishment concerned.[82]

Notwithstanding the above, Collective Bargaining Convention (No. 154) made remarkable improvements by including the whole public service (with the exception of the armed forces and the police) in the collective bargaining process.[83] The only condition is that special modalities of application can be fixed by national laws or regulations or national practice.[84] Also, the Labour Relations (Public Service) Convention (No 151) requires states to promote machinery for negotiation or such other methods that will allow representatives of public employees to participate in the determination of the terms and conditions of employment in the public service.[85]

Although as a justification for the exclusion of the 'armed forces' from trade union membership, it may be admitted that the susceptibility of danger of some targeted political and social activities emanating from their formation thereof, may hamper the attainment of the objects sought to be shielded in section 45 of the 1999 Constitution Federal Republic of Nigeria,[86] As regards the exclusion of certain classes of private and public officers, one cannot readily see, save for the perceived conflict of interest, how mere membership of an association whose primary objects must have been considered to be lawful prior to its due registration as a trade union constitute a danger sufficient to warrant derogation from the fundamental right to freedom of association guaranteed under section 40 of the 1999 Constitution. This assertion appears to be a truism to the extent that the registrar of trade unions is by virtue of section 7 (1)(b) and (d) of the Trade Unions Act, vested with the power to cancel any registration where "the principal purposes

[80] Trade Unions Act, s 11(1)(h).
[81] ibid s 11(i).
[82] ibid s 11(2).
[83] Convention No 154 (n 25), art 1(2).
[84] ibid art 1(3).
[85] Convention No 151 (n 37), art 7.
[86] Section 45 of the 1999 Constitution seeks to save 'any law that is reasonably justifiable in a democratic society (a) in the interest of defence, public safety, public order, public morality or public health; or (b) for the purpose of protecting the rights and freedom of other persons'.

for which the union is being carried on is a purpose other than that of regulating the terms and conditions of employment of workers". More so, any individual "engaged in acts calculated to disrupt the economy or [...] obstruct the smooth running of any essential service"[87] does so at the risk of a heavy financial penalty or a term of imprisonment or both".[88] These safeguards in terms of discretionary investiture and prescription of penalties are sufficient enough to keep extremist unions or organisations in check.[89] Not only is it unfair to seek to deprive a class of the Nigerian citizenry their constitutional right of association – for membership is no more than just that – merely because they are in the service of the state or community.[90]

C. Subject-matter of collective bargaining: what is negotiable?

The principle of free collective bargaining also implies that the parties have the right to negotiate collective agreements on all subjects of their choice.[91] In other words, the parties should be able to determine the subject matter and scope of negotiable issues.[92] Conventions No. 98, No. 151 and No. 154 and Recommendation No. 91 focus the content of collective bargaining on terms and conditions of work and employment and on the regulation of the relations between employers and workers and between organisations of employers and of workers.

The concept of working conditions used by the supervisory bodies is not limited to traditional working conditions (working time, overtime, rest periods, wages and so on.), but also covers "certain matters which are normally included in conditions of employment", such as promotions, transfers, dismissal without notice and so on.[93] This trend is in line with the modern tendency in industrialized countries to recognize "managerial" collective bargaining concerning procedures to resolve problems, such as staff reductions, changes in working hours and other matters which go beyond terms of employment in their strict sense.[94] The ILO Committee of Experts indicated that, "it would be contrary to the principles of

87 See section 1 (1) (a) of the Trade Disputes (Essential Services) Act 1976.

88 ibid s 1 (5).

89 Similar safeguards exist in other enactments, for example, the president is empowered by proclamation under section 1 (1) of the Trade Disputes (Essential Services) Act to proscribe such recalcitrant trade union. And what is more, the president pursuant to section 366 of the Criminal Code Act 1990 has similar power to proscribe any organisation under the Act, where the activities of such organisation constitute a danger to the security of the state or of its citizens.

90 Emiola (n 78) 415.

91 Okene, 'The Challenges of Collective Bargaining in Nigeria' (n 1) 88.

92 ibid 89.

93 ILO, 'General Survey Report' (n 76) para 250.

94 Gernigon and others (n 40) 39.

Convention No. 98 to exclude from collective bargaining certain issues such as those relating to conditions of employment" and "measures taken unilaterally by the authorities to restrict the scope of negotiable issues are often incompatible with the convention".[95]

Nevertheless, although the range of subjects which can be negotiated and their content is very broad, they are not absolute and need to be clearly related to conditions of work and employment or, in other words, matters which are primarily or essentially questions relating to conditions of employment.[96] Generally, the management representatives seek to define and limit the scope of collective bargaining in concrete terms".[97] They seek to establish a distinguishing line between management functions or management rights, otherwise conceptualized as "prerogatives", not subject to contractual rule-making and matters properly amenable to joint decision making.[98] The difficulty, however, lies in the general terms of these specific managerial issues which tend to overlap with the negotiable issues because they are ultimately two perspectives of a single set of interests which co-exist in the context of unity and variation. The common practice is to state as follows:

> The union undertakes not to interfere with the normal functions of management which give member companies of the Association the sole right and responsibility to conduct their business in such a manner as they consider fit and to engage, promote, demote, transfer, and terminate the service of any employee.[99]

The question must then be how do we balance the overlap against the agitations of union representatives that collective bargaining must remain a fluid and dynamic process? It is suggested that the determining factor be based on a 'proximity principle' in the sense that where these policies have important consequences on conditions of employment, that they should be the subject of collective bargaining. For analytical purposes, we can examine the nature of negotiable issues dealt with in collective bargaining under the following four[100] broad categories, *viz*:

95 ILO, 'General Survey Report' (n 76) para 265.
96 ILO, 'Digest of Decisions and Principles' 1996 (n 52) para 812.
97 Sylvia O Ebhoman, 'A Critical Examination of Collective Bargaining and its Role in Labour Relations in Nigeria' (LLM thesis Ahmadu Bello University 2016) 32.
98 ibid.
99 Tayo Fashoyin, *Industrial Relations in Nigeria* (Longman Nigeria Ltd 1992).
100 International Training Center of ILO: Bureau for Workers' Activities (Actrav) Course, 'Issues of Collective Bargaining' chp 3 <actrav-courses.itcilo.org> accessed 20 October 2021.

1. Wage Related Issues: These include issues like how basic wage rates are determined, cost of living adjustments, wage differentials, overtime rates, wage adjustments and so on.
2. Supplementary Economic Benefits: These include issues as pension plans, paid vacations, paid holidays, health insurance plans, dismissal plans, supplementary unemployment benefits and so on.
3. Institutional Issues: These consists of rights and duties of employers, employees and unions, including union security, check off procedures, hour of work, quality of work-life program and so on.
4. Administrative Issues: These include issues such as seniority, employee discipline and discharge procedure, employee health and safety, technological changes, work rules, job security and training, attendance, leave and so on.

In Nigeria, the position is that the scope of negotiable issues in collective bargaining is subject to certain restrictions. In the public sector, negotiable issues are spelt out in the National Public Service Negotiating Council (NPSNC). Many of the substantive issues which are within the scope of the NPSNC are made either by legislative or executive acts or through political commission periodically set up by government as employer of labour.[101] The issues as Agomo notes are threefold: namely, negotiation on all matters affecting the conditions of service of all civil servants; advising government when necessary on how to harness ideas and experience of civil servants for improved productivity; reviewing the general conditions of civil servants.[102] In practice, however, as Fashoyin has pointed out, many items of conditions of service such as salary, leave entitlements, minimum wage, pensions and car loan are excluded from negotiation.[103] On the other hand, negotiable issues in the private sector are contained in the procedural agreement which contains guidelines on the standards, methods and levels to be followed by the negotiating parties.[104] It contains the subjects for negotiation at each bargaining level and also clarifies issues of management prerogatives on which negotiation is not allowed.[105] Procedural agreements accord recognition to the unions and usually affirm principle of co-operation and peaceful relations between trade unions and the employers.[106] Additionally, the bargaining unit for the different categories of employment as well as the machinery for negotiations are

[101] Olulu and Udeorah (n 5) 66.
[102] Chioma K Agomo, *Nigeria in International Encyclopaedia of Comparative Labour Law and Industrial Relations* (Kluwer Law International 2000) 249.
[103] Fashoyin (n 99) 165.
[104] Okene,'The Challenges to Collective Bargaining in Nigeria' (n 1) 86.
[105] Olulu and Udeorah (n 5) 66.
[106] Agomo (n 102) 249.

included in the procedural agreement.[107] The custom of explicitly laying down in the procedural agreement definite terms and conditions which are subject to negotiation, to the exclusion of other matters for discussion and consultation, does not align with the standard prescriptions of ILO, as it permits management to claim prerogative power over certain matters relating to the promotion and discipline of employees.[108]

D. THE LEVELS OF COLLECTIVE BARGAINING

Collective bargaining takes place at several organisational levels. There is no generally accepted best level for collective bargaining.[109] The appropriate level or levels for bargaining depend on the strength, interests, objectives and priorities of the parties concerned, as well as the structure of the trade union movement, employers' organisation and traditional patterns of industrial relations.[110] The three basic levels at which collective bargaining are usually conducted are the enterprise level, the industry level and the plant or individual workplace level.[111] At the enterprise level, collective bargaining involves an employer on the one hand and the trade union that caters for the interest of his employees on the other. Collective bargaining at the industry level normally takes place between an industrial union and an industry-based employers association. The lowest level at which collective bargaining may take place is at the workplace itself.[112] In terms of ILO Law, the level at which collective bargaining between the employer and her/his employees or their respective representatives is to be effected is generally a matter to be decided upon by the parties themselves. The ILO Collective Bargaining Recommendation No 163 provides that:

> Measures adapted to national conditions should be taken, if necessary, so that collective bargaining is possible at any level whatsoever, including that of the establishment, the undertaking, the branch activity, the industry, or the regional or national levels.[113]

[107] ibid.

[108] Issues bordering on promotion, discipline and transfer amongst others have traditionally been regulated by the Civil Service Rules, and this undoubtedly whittles down the voluntariness of collective bargaining in the public sector.

[109] Okene, 'The Challenges of Collective Bargaining in Nigeria' (n 1) 86.

[110] Greg Bamber, *Comparative Labour Law and Industrial Relations in Industrialized Market Economies* (Kluwer Law International 1998) 414.

[111] Okene, 'The Challenges of Collective Bargaining in Nigeria' (n 1) 86.

[112] ibid.

[113] Recommendation No 163 (n 37), para 4 (1).

In Nigeria, collective bargaining in the private sector takes place at four levels such as: (1) the industry level, which is between an industrial union and an industrial employers' association: (2) the company level, which is between an industrial union and individual employers: (3) the branch or enterprise level, which is usually between the branch of the industrial union and the company management, and (4) the plant level, between the plant unit of the branch union and the plant management.[114] In the public sector, the framework for collective bargaining is through the NPSNC. As Agomo pointed out, the NPSNC envisages collective bargaining in the public Sector to take place at three levels such as the Federal level, the State level and the Ministerial level. Bargaining at the Federal level is further split into three categories, those representing senior staff on grade levels 10-14, junior staff on grade levels 01-06, and technical staff.[115] In practice, successive governments have had to make use of *ad-hoc* commissions[116] in the determination of wages and conditions of service of public sector workers.[117]

Although the parties to collective bargaining in the private sector may voluntarily select the level at which to bargain, in the public sector the Nigerian government unilaterally decides for her workers, as they are subjected to decisions by the *ad-hoc* commissions. This means, in effect, that there is no level of bargaining to choose from[118] which is contrary to the CFA ruling that "the determination of bargaining level is essentially a matter to be left to the discretion of the parties and, consequently, the level of negotiation should not be imposed by law, by decision of the administrative authority."[119] In this wise also, the neutrality and independence of the NPSNC remains doubtful because many of the substantive issues which are within the domain of the NPSNC are decreed either by executive or legislative acts or via political body like commission periodically created by government as employer of labour. The role of NPSNC Nigeria is totally irrelevant because of the influence and role of other government agencies.[120] These

[114] Fashoyin (n 99) 119.

[115] Agomo (n 102) 249.

[116] Some of these commissions include: the Morgan Commission of 1963-64, the Wages and Salaries Review Commission (Adebo Commission) of 1970-71, Public Service Review Commission (Udoji Commission) of 1972-74, Pension Reform Commission of 2004, Wages and Salaries Review Commission (Shonekan Commission) of 2005-2006, Onosode Commission (for parastatals) of 1981, Adamolekun Commission (for Polytechnics) of 1981, Ukandi Damachi Commission of 1990, Minimum Wage Commission of 1999, Ufot Ekaette Presidential Committee on Monetization of Fringe Benefits in the Public Service of 2002, Onosode Commission (for universities) of 2009.

[117] Agomo (n 102) 249.

[118] Okene, 'The Challenges of Collective Bargaining in Nigeria' (n 1) 88.

[119] ILO, 'Digest of Decisions and Principles' 2006 (n 56) para 988.

[120] Ugbomhe Ugbomhe and NG Osagie, 'Collective Bargaining in Nigeria: Issues, Challenges and Hopes' (2019) 7 Journal of Human Resources Management and Labour Studies 20.

developments have undermined the relevance of collective bargaining in the public sector.[121]

It is submitted that Nigeria is in breach of ILO standards for failing to allow workers in the public sector to bargain at an appropriate level.[122] This is despite the provision of three levels of bargaining in the public sector through the NPSNC as discussed above. In India, for example, the position is remarkably different. Collective bargaining takes place at various levels. The choice of level appears to vary according to the category of workers. In the private sector, collective bargaining takes place at plant level between the management of the plant and an enterprise-based union. In public sector enterprises, bargaining takes place between centralised trade union federations and the State (as employer) at industry and, or national level. Central and State government employees in the service sector bargain at national and or regional level through affiliated unions.[123] There is the need therefore to change the position in Nigeria so that both private and public sector workers can freely choose the level at which they wish to bargain. This will bring Nigerian law into conformity with international labour standards.[124]

III. Governing Principles of ILO standards

A. The principle of free and voluntary negotiation

The framework within which collective bargaining must take place if it is to be viable and effective is based on the principle of the independence and autonomy of the parties and the free and voluntary nature of the negotiations; it requires the minimum possible level of interference by the public authorities in bipartite negotiations and gives primacy to employers and their organisations and workers' organisations as the parties to the bargaining.[125] This principle is embodied in the Right to Organise and Collective Bargaining Convention No 98, which was adopted in 1949, and which since has achieved near-universal acceptance: as of September 2021 the number of member States having ratified it stood at 168,[126]

[121] FC Anyim and others, 'Collective Bargaining Dynamics in the Nigerian Public and Private Sectors' (2011) 1 Australian Journal of Business and Management Research 63.

[122] Okene, 'The Challenges of Collective Bargaining in Nigeria' (n 1) 88.

[123] Debashish Battercherjee, 'Organised Labour and Economic Liberalisation in India: Past, Present and Future' (1999) International Institute for Labour Studies Research Paper 105/1999, 1-62 <https://www.google.com/url?sa=t&source=web&rct=j&url=https://library.fes.de/pdf-files/gurn/00166.pdf&ved=2ahUKEwiC1Naz6eXzAhXeA2MBHTZSCEsQFnoECAQQAQ&usg=AOvVaw1HK5FXvVy0HQOxtlHF-NWs> accessed 25 October 2021.

[124] Okene, 'The Challenges of Collective Bargaining in Nigeria' (n 1) 88.

[125] Gernigon and others (n 40) 34.

[126] ' Ratifications of Right to Organise and Collective Bargaining Convention No 98' (*International Labour Organisations Normlex,* 21 September 2021)

which demonstrates the force of the principles involved in the majority of countries. Convention No 98 does not contain a definition of collective agreements, but outlines their fundamental aspects in Article 4:

> Measures appropriate to national conditions shall be taken, where necessary, to encourage and promote the full development and utilisation of machinery for voluntary negotiation between employers or employers' organisations and workers' organisations, with a view to the regulation of terms and conditions of employment by means of collective agreements.

It is pertinent to emphasise that, for workers' organisations to be able to fulfil their purpose of "furthering and defending the interests of workers" through collective bargaining, they have to be independent[127] and must be able to organise their activities without any interference by the public authorities which would restrict this right or impede the lawful exercise thereof.[128] Moreover, they must not be "under the control of employers or employers' organisations".[129] Under "voluntarism", employers and unions have reasonable latitude to determine their own affairs within a framework established by the state.[130] As Fashoyin has pointed out, "this doctrine emphasises the freedom of labour and management to determine as much as possible the conditions under which workers will work, as well as other issues of labour relations".[131] It is often based on the theory that those closest to industry are in the best position to solve any problems arising from labour and management relations: in short it is a theory of industrial self-governance.[132] Furthermore, the principle of voluntary collective bargaining was pursued in the belief that it was better suited for the sustenance of industrial peace and harmony than the interventionist approach.[133]

Similarly, the principle of voluntarism in negotiation transcends to include machinery which supports bargaining such as the provision of information, consultation, mediation, arbitration. The CFA has established that the bodies appointed for the settlement of disputes between the parties to collective bargaining should be independent, and recourse to these bodies should be on a

<https://www.ilo.org/dyn/normlex/en/f?p=NORMLEXPUB:11300:0::NO::p11300_instrument_id:312243> accessed 21 September 2021.

127 See Convention No 87 (n 23), art 3 (1).

128 ibid art 3(2).

129 See Convention No 98 (n 24), art 2(1) and (2).

130 AS Egbo, 'Contemporary Issues in Public Sector Collective Bargaining' in Tayo Fashoyin (eds), *Collective Bargaining in the Public Sector in Nigeria* (MacMillan1987) 24.

131 Fashoyin (n 96) 97.

132 ibid.

133 ibid.

voluntary basis.[134] The supervisory bodies admit conciliation and mediation which are voluntary or imposed by law, if they are within reasonable time limits[135] — as well as voluntary arbitration — in accordance with the provisions of Recommendation No 92 which indicates that "[p]rovision should be made to enable the procedure to be set in motion, either on the initiative of any of the parties to the dispute or ex officio by the voluntary conciliation authority".[136]

Drawing from the above, it follows that the obligation to promote collective bargaining excludes recourse to measures of compulsion. During the preparatory work for Convention No 154, the Committee on Collective Bargaining agreed upon an interpretation of the term "promotion" (of collective bargaining) in the sense that it "should not be capable of being interpreted in a manner suggesting an obligation for the State to intervene to impose collective bargaining", thereby allaying the fear expressed by the employer members that the text of the Convention could imply the obligation for the State to take compulsory measures.[137] The Committee on Freedom of Association, following this line of reasoning, has indicated that:

> Collective bargaining, if it is to be effective, must assume a voluntary quality and not entail recourse to measures of compulsion which would alter the voluntary nature of such bargaining.[138]

As a former British colonial territory, Nigeria's industrial relations system was fashioned in line with the British industrial relations system whose "main feature is the voluntary machinery which has grown over a wide area of employment for industry-wide collective bargaining between employers' associations and trade unions over terms and conditions of employment."[139] The policy was as an overt expression of the government's perception that it was better to leave both employers and employees free to determine and regulate their relations as best as they could.[140] Okotie -Eboh, Minister of Labour, stated this policy thus:

[134] ILO, 'Digest of Decisions and Principles' 1996 (n 52) paras 858–859.
[135] ibid paras 502–504.
[136] Recommendation No 92 (n 37), Para 3.
[137] International Labour Organisation, *Record of Proceedings* (1981) ILC67, 22.
[138] ILO, 'Digest of Decisions and Principles' 1996 (n 52) para 845.
[139] See Royal Commission: 'Written Evidence of the Ministry of Labour' cited in H. Clegg, *The System of Industrial Relations in Great Britain* (Basil Blackwell 1976) 200.
[140] AA Tajudeen and OK Kehinde, 'Government Public Policies and the Dynamics of Employment Relations in Developing Countries: The Experience of Nigeria' (2007) 4 Pakistan Journal of Social Science 761.

> We have followed in Nigeria the voluntary principles which are so important an element in industrial relations in the United Kingdom.... Compulsory methods might occasionally produce a better economic or political result, but labour-management must, I think, find greater possibilities of mutual harmony where results have been voluntarily arrived at by free discussion between the two parties. We in Nigeria, at any rate, are pinning our faith on voluntary methods.[141]

Over the years, the successive governments have been fully intervening in industrial relations. The interventionist role can be seen to be the result of the proliferated incidence of military usurpation and administration in Nigeria with several of its labour decrees being weighted heavily against labour. It would be apposite in the author's view that the protection via the collective bargaining mechanism accorded to both parties, unjustly tilts towards the management. For instance, the right to collective bargaining is restricted by the requirement for government approval. Although, in theory it is settled law that failure to accord recognition to trade union during collective bargaining is unlawful, however, it is the position of the law that every terms of collective agreement must be confirmed in an order of the minister of labour as a precondition for its enforceability on the employers and workers to whom they relate.[142] This interventionist approach as opposed to "voluntarism," whittles down the latitude of employers and unions to reasonably determine their own affairs within a framework established by the state.

Proceeding from a similar approach, the practice of routing disputes for settlement through the minister of labour is reminiscent of the government's interventionist policies. The Trade Disputes Act does not allow workers and trade unions to take their disputes directly to the arbitral bodies. Only the Minister of Labour alone is empowered to make such a decision.[143] For example, it is he who appoints a fit person as a conciliator for the purpose of effecting a settlement of a trade dispute.[144] The Industrial Arbitration Panel (IAP) can only act upon a case referred to it by the Minister.[145] Moreover, in the case of the lAP any award is communicated to the Minister alone and not the parties affected.[146] This discretionary investiture to refer disputes to the arbitral bodies vis-à-vis the statutory mandate to appoint the conciliator as well as members of the Board of Inquiry and the Arbitration Tribunal, raises fears over the neutrality and

[141] International Labour Organisation, *Record of Proceedings* (1958) ILC88, 33.
[142] Trade Disputes Act, s 3(1).
[143] ibid ss 8, 9, 17 and 33.
[144] ibid s 8.
[145] ibid s 9.
[146] ibid s 13.

independence of this procedural arrangement and its susceptibility to extrinsic influence emanating from the political cadre of the economy. This undoubtedly defeats the objective of the machinery for the settlement of trade disputes, which is to temporarily suspend the right to strike and provide an adequate, impartial and speedy resolution of disputes.[147] This is part of the inbuilt bottlenecks which are capable of slowing down the process. The ILO does not support cumbersome and complicated dispute resolution processes which tend to frustrate the right to strike. In ILO's view, "such machinery must have the sole purpose of facilitating bargaining: it should not be so complex or slow that a lawful strike becomes impossible in practice or loses its effectiveness".[148]

Going further, in the public sector, for instance, the government has arrogated to itself the role which both employers and employees supposed to perform in industrial relations.[149] As a state authority, government set up machinery i.e. councils to negotiate for salary increase and other conditions of service in the public sector.[150] Imafidon correlated the current position of government in wage fixing when he advanced the argument that collective bargaining has been relegated to the background in Nigeria because government resorted to creating wage tribunal as a mechanism of fixing and reviewing wage.[151] Lending credence to this view, several writers have opined that the use of ad-hoc commission in addressing workers' demand such as wage determination and other term and conditions of employment is unilateral and undemocratic as it violates good industrial democratic principles.[152] Nigeria determination of minimum wages has always been carried out without any effective tripartite collective bargaining, the latest being the new minimum wage effectuated by the current regime of Muhammadu Buhari in 2019. This development not only makes it antithetical to democratic value, but has also undermined the importance of collective bargaining in Nigeria public sector.[153]

Overall, although the government's policies on labour relations are anchored on what it called "limited intervention guided democracy", the evidence suggests otherwise.[154] Rather, as has been seen, government's policies and the dynamics of labour relations demonstrate that what obtains is unguided

147 OVC Okene, 'Mechanisms for the Resolution of Labour Disputes in Nigeria: A Critique' (2010) 3 Kogi State University Bi-annual Journal of Public Law 151.

148 ILO, 'General Report Survey' 1994 (n 76) para 171.

149 Ugbomhe and Osagie (n 120) 30.

150 ibid.

151 Tongo Imafidon and Osabuohien Evans, 'Emergent and Recurrent Issues in Contemporary Industrial Relations: Pathways for Converging Employment Relationships' (2007) 4 Journal of Management and Enterprise Development 43.

152 Anyim and others (n 121) 63.

153 Ugbomhe and Osagie (n 120) 30.

154 OVC Okene, 'Nigeria's Labour and Industrial Relations Policy: From Voluntarism to Interventionism — Some Reflections' (2012) 4 Port Harcourt Law Journal 240.

authoritarianism and reckless intervention in labour relations.[155] Through its policies and laws the government has seriously infringed the rights of Nigerian workers. In this wise, it is thus clear that the government interventionist policy indicated a systematic approach that was largely repressive of labour rights, and in particular pointed to the state's high-handedness as far as workers are concerned.[156]

B. The principle of bargaining in good faith

In order for collective bargaining to be workable, it should be conducted in good faith by the parties to the negotiation. Having been duly recognized, it follows that a trade union would expect the employer to be willing to enter into genuine negotiations with it.[157] Prospective as it may seem, the reality is that most employers shy away from negotiating voluntarily and faithfully.[158] Consequently, the need to foist on employers not only an obligation to bargain collectively, but also to do so in good faith becomes apposite. In the preparatory work for Convention No 154, it was recognized that collective bargaining could only function effectively if it was conducted in good faith by both parties; but as good faith cannot be imposed by law, it "could only be achieved as a result of the voluntary and persistent efforts of both parties".[159] The CFA, in addition to drawing attention to the importance that it attaches to the obligation to negotiate in good faith, has established four guiding principles about what good faith entails. According to the Committee, "good faith" implies:

1. Making every effort to reach an agreement (or settlement as the case may be);
2. Conducting genuine and constructive negotiations;
3. Avoiding unjustified delays; and
4. Comply with the agreements which are concluded and applying them in good faith. [160]

In Nigeria the "obligation to bargain in good faith" is not expressly provided for in any of the laws dealing with employment matters, and this appears to be one of the impediments to collective bargaining in Nigeria. In the public

155 Tajudeen and Kehinde (n 140) 761.
156 S Okodudu and BK Girigiri, 'The State and Labour Militancy in Nigeria' (1998) 3 Pan-African Social Science Review 34.
157 Okene, 'The Challenges of Collective Bargaining in Nigeria' (n 1) 84.
158 ibid.
159 ILO, 'Record of Proceedings' 1981 (n 141) 22.
160 ILO, 'Digest of Decisions and Principles' 1996 (n 52) paras 814–818.

sector, for example, the lack of good faith bargaining is attributed to the limited authority of civil servants who represent government on the bargaining table, and as such, a practical implication of this is the unduly long process it takes to give final approval to decisions reached at negotiations.[161] In this regard, Okene notes that there exists a chain of decision-making processes which may originate from the negotiating table but goes on to the various governmental agencies up to the highest level in the political authority.[162] Government officials lack the authority to firmly and in good faith commit the state at negotiations with the workers or their representative union.[163] A practical implication of this is the unduly long process it takes to give final approval to decisions reached at negotiations.[164]

This practice undoubtedly contravenes the process of conducting genuine and constructive negotiations and to conclude agreements in good faith as required by the ILO. As Fashoyin noted, "the dichotomy between those undertaking negotiation and the deciding authorities is such to make it appear to the workers that it is wilfully calculated to frustrate their demands".[165] It is therefore of utilitarian value that Nigeria should provide for the duty to bargain in good faith, both in the private and public sectors to effectively promote the practice of collective bargaining. Lending credence to this view, Okene rightly points out that "there is no point engaging in collective bargaining, if the parties cannot negotiate with an honest intention of reaching an agreement which they intend to bind them".[166]

C. Principle of Enforcement of Collective Agreements

In the ILO's instruments, collective bargaining is deemed to be the activity or process leading up to the conclusion of a collective agreement.[167] In Recommendation No 91, Paragraph 2, collective agreements are defined as:

> All agreements in writing regarding working conditions and terms of employment concluded between an employer, a group of employers or one or more employers' organisations, on the one hand, and one or more representative workers' organisations, or, in

[161] Sola Fajana, *Industrial Relations in Nigeria: Theory and Practice* (Labofin and Company 2000) 274.
[162] Okene, 'The Challenges of Collective Bargaining in Nigeria' (n 1) 85.
[163] ibid.
[164] Fajana (n 161) 274.
[165] Tayo Fashoyin, 'Collective Bargaining in the Public Sector: Retrospect and Prospects' in Tayo Fashoyin (eds), *Collective Bargaining in the Public Sector* (Macmillan press1987) 11.
[166] Okene, 'The Challenges of Collective Bargaining in Nigeria' (n 1) 86.
[167] Gernigon and others (n 40) 35.

> the absence of such organisations, the representatives of the workers duly elected and authorised by them in accordance with national laws and regulations, on the other.

The Recommendation No. 91 goes on to state that collective agreements should bind the signatories thereto and those on whose behalf the agreement is concluded[168] and that stipulations in such contracts of employment which are contrary to a collective agreement should be regarded as null and void and automatically replaced by the corresponding stipulations of the collective agreement.[169] Stipulations in contracts of employment which are more favourable to the workers than those prescribed by a collective agreement should not be regarded as contrary to the collective agreement.[170] It set out the binding nature of collective agreements and their precedence over individual contracts of employment, although recognizing the stipulations of individual contracts of employment which are more favourable for workers.

Comparatively, under the Nigerian labour law, there is no presumption of intention about the binding force of a collective agreement between the parties thereto. The nearest it has gone in attaching legal enforceability to a collective agreement is in the provision of Section 3(1) of the Trade Disputes Act which stipulates expressly that parties to a collective agreement are expected to deposit with the minister of labour and productivity at least three copies of the agreement within 30 days of its execution, and when such deposit is made the minister may by order make the agreement or part thereof binding on the parties to whom it relates. The Nigerian Courts have taken the common law position that collective agreement is merely "a gentleman agreement and is binding only in honour and not enforceable". In *Union Bank of Nigeria v Edet,*[171] the employee's contention that her termination flouted the collective agreement was rejected. It was held that collective agreements, except where they have been adopted as forming part of the terms of employment, are not intended to give or capable of giving an individual employee the right to institute an action for breach of any collective agreement, nor is it intended to complement the employee's contract of service.[172] It was noted that:

> Collective agreements are not intended or capable of giving individual employee a right to litigate over an alleged breach of their terms as may be conceived by them to have affected their interest,

168 Recommendation No 91 (n 37), para 3(1).
169 ibid para 3(2).
170 ibid para 3(3).
171 *Union Bank of Nigeria v Edet* [1993] 4 NWLR (Pt 287) [298]–[299].
172 ibid [288] (Uwaifo JCA).

> nor are they meant to supplant or even supplement their contract of service. In other words, failure to act in strict compliance with collective labour agreement is not justiciable.[173]

It may be argued that the courts' refusal to enforce collective agreements is based on the privity of contract,[174] as most collective agreements are usually between the employers on one part and trade unions on the other. An individual employee seeking to benefit from it is not party to it.[175] In *Afribank (Nig) Plc v Osisanya*[176] Amaizu JCA held that the dismissal procedure contained in the collective agreement was not binding on the employer as the agreement was not justiciable. In *ACB Plc v Nwodika*,[177] Ubaezonu JCA outlined factors which may determine whether a collective agreement is binding on individual employees and employers: namely: its incorporation in the contract of service, if any, the pleadings and evidence before the court or the parties' conduct.

From the above, it is clear that one of the challenges that plagues the practice of collective bargaining in Nigeria is that of non-observance of collective agreement which is the finished product of collective bargaining. Paradoxically, while it may be adduced that the issue of enforceability has statutory backing under section 3(2) of the Trade Disputes Act 2004, although not full-fledged in the true sense of ILO's standard prescriptions on "voluntarism" in negotiation, the issue of judicial recognition of such collective agreements has always become revolving challenge in Nigeria. It seems lamentable that agreements wrapped up through collective bargaining cannot be readily enforced. Perhaps, it may be possible to enforce collective agreements in Nigeria under the Constitution. A new provision – section 254C (2) – was introduced into the Constitution in 2010 empowering the National Industrial Court of Nigeria (NICN) exclusively to apply any ratified international treaty relating to labour and industrial relations. For clarity, section 254 C (2) provides thus:

> Notwithstanding anything to the contrary in the Constitution, the National Industrial Court shall have the jurisdiction and power to deal with any matter connected with or pertaining to the application of any international convention, treaty or protocol of which Nigeria has ratified relating to labour, employment, workplace, industrial relations or matters connected therewith.

[173] ibid [298].
[174] ibid.
[175] ibid.
[176] *Afribank (Nig) Plc v Osisanya* [2000] 1NWLR (Pt. 642) 598.
[177] *ACB Plc v Nwodika* [1996] 4 NWLR (Pt 443) 470, 485. See also *B.P.E. v Dangote Cement Plc* [2020] 5 NWLR (Pt. 1717) 322.

It is important to note that Nigeria has ratified ILO Convention 98 (concerning collective bargaining). Thus the proviso in the section which states *"notwithstanding anything to the contrary in the Constitution"* now appears to vest the NIC with the right to apply international labour conventions ratified by Nigeria. This is clearly derogation from section 12(1) of the same Constitution and which is to the effect that a treaty shall not have the force of law in Nigeria except same has been enacted into law by the National Assembly. The implication is that the NICN could enforce that collective agreement through section 254C (2) since Nigeria has ratified Convention 98. Domestication is not Required before enforcement by the provision. This is buttressed by section 7(6) of the National Industrial Court Act further provides a legal ground for the contention that non domesticated conventions can be applied as examples of international best practices. The section provides that:

> The court shall, in exercising its jurisdiction or any of the powers conferred upon it by this Act or any other enactment or law, have due regard to good or international best practices in labour or industrial relations and what amounts to good or international best practice in labour or industrial relations shall be a question of fact.

Commenting on this position, Hon. Justice B.B Kanyip rightly opined that:

> Section 7(6) of the National Industrial Court provides an avenue for Nigeria, as a member of the international community, and as a member of International Labour Organisation, to take advantage of international labour jurisprudence in the resolution of domestic issues.[178]

Therefore, a restrictive interpretation of the Constitution should not be used to hinder the implementation of Nigeria's voluntary membership and ratification of international obligations, especially as regards the Conventions and Recommendations of the ILO.

There is no doubt that the interventionist policy under section 3 (2) of the Act wherein the Minister wields such wide discretionary powers is subject to abuse. The Minister may in dereliction of his duty or in the exercise of the latitude of his discretionary investitures under the Act refuse to make an order confirming the terms of a duly concluded collective agreement. Indeed, as Okene rightly points

[178] Benedict B Kanyip, 'Current Issues in Labour Dispute Resolution in Nigeria' (All Judges Conference, Abuja, November 20).

out, "the Minister will never make such an order especially where the interests of the government whom he represents will be affected by the order".[179] Without the enforceability of collective agreements collective bargaining is but a mere vain exercise and cannot be effective. As aforementioned, the Committee on Freedom of Association has ruled that all collective bargaining agreement should be binding on the parties. The Committee on Freedom of Association has also ruled that making the validity of collective agreements signed by the parties subject to the approval of these agreements by the authorities is contrary to the principles of collective bargaining and of Convention No 98.

It is submitted therefore that Nigerian legal framework must expressly provide that once agreements are concluded by the parties thereto they become readily enforceable without further ado. Nigeria can take in tow the labour statutes in some African countries which contain comprehensive provisions regarding the enforceability of collective agreements. For instance, labour statutes in Ghana,[180] Kenya,[181] Zambia,[182] and South Africa[183] (which are of common law jurisdiction like Nigeria) expressly provide that collective agreements relating to employment and labour are binding and enforceable. The implication is that the courts in those countries will enforce any collective agreement concluded between an employer and his employees without considering the common law position as the provisions of a statute always prevail over the common law.[184]

[179] Okene, 'The Challenges of Collective Bargaining in Nigeria' (n 1) 97.

[180] Section 105 (2) of the Ghanaian Labour Act (No. 651 of 2003) states that collective agreement between employees and an employer is viewed as terms of the employment contract between each employee and his employer.

[181] In Kenya, s 59 (1) of the Labour Relations Act (No 14 of 2007) stipulates that every collective agreement relating to employment and labour binds all employees and their employers. Section 59 (3) also states that every collective agreement should be incorporated into the employment contract of an employee. Furthermore, section 59 (5) states that upon registration of the collective agreement, it becomes enforceable.

[182] Section 71(3) (c) of the Zambian Industrial and Labour Relations Act (No. 27 of 1993, Cap 269 of the Laws of Zambia) states that once a collective agreement has been accepted by the Minister, it becomes binding between the employer and employee or between the parties.

[183] Section 23 of the South African Labour Relations Act (No. 66 of 1995) stipulates that every collective agreement relating to employment and labour binds not only the parties to it, but other persons to which the collective agreement applies. Also, section 199 states that an employment contract entered into before or after a collective agreement may not allow an employer to pay his workers remuneration less than what is stipulated in the collective agreement. It further provides that any contract that purports to waive any collective agreement is invalid.

[184] See s 1(3) of the 1999 Constitution.

IV. Comparative Analysis of Collective Bargaining in Other Jurisdictions: The United Kingdom And South Africa Example

A. United kingdom

It is rather safe to begin by stating that Nigeria was once a British colony and most of her laws were derived from the common law provisions. Indeed, a peep into collective bargaining as practiced in the UK becomes apposite. Perhaps no other country in recent years has witnessed greater change in its collective bargaining framework than the UK. The English Trade Union and Labour Relations (consolidation) Act, 1992 brought to light, amongst other things, the seamless operation of the collective bargaining mechanism in the UK. Although the Trade Union and Labour Relations (consolidation) Act does not provide for the obligation to bargain but merely facilitates collective bargaining, leaving the rest to the parties involved; It however imposes a duty on the employer to disclose to a representative trade union all relevant information that will enable effective collective bargaining thus indirectly adopting the duty to bargain into its framework.[185] Unlike in Nigeria where the scope of negotiable issues is subject in collective bargaining is subject to certain restrictions, the Trade Union and Labour Relations (consolidation) Act provides for a wide range of negotiable matters covered in the collective bargaining process.[186]

Furthermore, it appears that Nigeria have been left behind because, there have been a paradigm shift through legislation from the common law position on the doctrine of unenforceability of collective agreements. Today in the UK, the doctrine that a third party cannot enforce a contract has ceased to be the law. A third party can now enforce a contract in two situations; firstly, if the third party is mentioned in the contract as the person authorized to enforce it and secondly if the contract purports to confer a benefit on the third party. Presently, collective agreements are enforceable in the UK once the parties include in the agreement, a provision that it would be legally binding on the parties. Under the Trade Union and Labour Relations (consolidation) Act, a collective agreement is presumed enforceable where it is in writing and provides expressly that the agreement is legally binding on the parties thereto.[187] Thus, the doctrine of privity of contract no longer weighs down collective agreements in England and such agreements become automatically enforceable between the parties if they are reduced into

[185] Trade Union and Labour Relations (Consolidation) Act, s 181.
[186] ibid s 178(2).
[187] ibid s 179.

writing and are stipulated to be legally binding. Notwithstanding that the doctrine has been buried in the UK from where it came to Nigeria; the Nigerian law makers and surprisingly the court, rather than build on this progressive assertion that a collective agreement reduced into writing and agreed upon is absolutely, legally binding and enforceable, held in plethora of cases as discussed above at Section III. C., that whether or not a collective agreement is binding on individual employees is dependent on its incorporation in the contract of service. Although traces of progress and divergence can be seen under the Constitution (Third Alteration) Act, however, the judicial emancipation of Nigerian laws from these vestiges of common law has been sluggish, and their traces and influence are very much evident in the jurisprudence of labour and industrial relations.

B. SOUTH AFRICA

The South African legal frameworks, the Constitution and its labour relations frameworks are amongst the most progressive institutions in the world.[188] Its Constitution stands apart in Africa having expressly entrenched the right of workers[189] and employers[190] to form trade unions and employers' organisations, guaranteeing the right of trade unions, employers' organisations and employers to engage in collective bargaining.[191] In terms of its judicial approach, the countries labour frameworks seek to fulfil South Africa's obligations as Member State of the ILO.[192] In cognisance and furtherance of this purpose, judges in South Africa also establish jurisprudential principles based on both ratified and non–ratified international labour standards.[193] The reason seems not to be far-fetched. Unlike Nigeria which is a dualist state,[194] in South Africa, a dualist approach is used in dealing with treaties and a monist-like approach is used for international customary international law.[195] In this regard, section 233 of the Constitution of the Republic of South Africa, 1996, provides to the effect that "every court must

[188] Tavonga J Zvogbo, 'Collective Bargaining and Collective Agreements in Africa: Comparative Reflections on SADC' (2019) International Training Center of ILO Working Paper 11/2019, 21 <https://www.google.com/url?sa=t&source=web&rct=j&url=https://www.itcilo.org/sites/default/files/inline-files/WP%252011_TJ%2520Zvobgo.pdf&ved=2ahUKEwiWxd_o64_2AhU2_7sIHcfaAWQQFnoECAQQAQ&usg=AOvVaw1m92tEhTpOlhuo85_4Qpuw> accessed 17 February 2022.

[189] Constitution of the Republic of South Africa, Act 108 of 1996, s 23 (2).

[190] ibid s 23 (3).

[191] ibid s 23 (5).

[192] See *National Union of Metal Workers of South Africa and Others v Bader Bop (Pty) Ltd and Another* [2003] 2 BCLR 182.

[193] See *Modise and Others v Steve's Spar Blackhealth* [2000] 5 BLLR 496.

[194] Constitution Federal Republic of Nigeria, 1999, s 12(1); See also Abacha and Others v Fawehinmi [2000] 6 NWLR 228.

[195] Constitution of the Republic of South Africa, ss 231 (4) and 232.

prefer any reasonable interpretation of the legislation that is consistent with international law" when interpreting any legislation but must consider international law when interpreting the Constitution's Bill of Rights. No doubt, any derogation from this constitutional prerogative by any court within South Africa would constitute sufficient grounds for review and appeal.

Furthermore, its liberal Labour Relations Act[196] (LRA) was enacted with the purpose of creating conditions for workers to act collectively to bargain with their employers effectively. Just like in the UK, the LRA does not provide for the duty to bargain but merely facilitates collective bargaining. In that regard, it imposes a duty on the employer to disclose to a representative trade union all relevant information that will enable effective collective bargaining thus indirectly adopting the duty to bargain into its framework.[197] Additionally, the LRA gives effect to the freedom to bargain collectively by providing the institutional infrastructure for voluntary collective bargaining at sector level and for the binding nature of collective agreements. The concern that voluntarism may allow employers to refuse to bargain at all is met to some extent by the organisational rights accorded to trade unions in Chapter III of the LRA and the provision of a statutory dispute resolution procedure. The LRA's approach is to provide the organisational infrastructure for union organisation at the workplace and to provide a conciliation procedure to resolve interest disputes irrespective of whether the trade union is recognised.[198]

Collective bargaining in South Africa much like in Nigeria, takes places at several levels. A distinction in South Africa can however, be seen between single-employer bargaining (branch, company or corporate level) and multi-employer bargaining (more than one employer represented by employers' organisation),[199] with the latter taking place in the form of bargaining councils. One key feature of multi-employer bargaining arrangements is that the agreements reached will be extended to non–parties, that is, to employers and employees who are not members of the organisations that negotiated the agreement.[200]

In South Africa, the practice and procedures of enforcement of collective agreements are entirely different from that of Nigeria in the sense that collective agreements are enforced as a matter of course by the parties, provided that they

196 No 66 of 1995.

197 Labour Relations Act, s 16 (3).

198 Halton Cheadle, 'Collective Bargaining and the LRA' (2005) 9 Law, Democracy and Development Journal 148.

199 Shane Godfrey, 'Multi-employer Collective Bargaining in South Africa' (2018) ILO Condition of Work and Employment Series Paper 97/2018, 1 <https://labordoc.ilo.org/permalink/41ILO_INST/8s7mv9/alma994995393302676> accessed 18 February 2022.

200 ibid.

are entered into or made in writing.[201] The collective agreement when decided upon, has the effect of altering the terms of any contract or employment relationship between an employee and an employer who are both bound by the collective agreement.[202] The LRA generally allows collective agreements to take precedence over its own provisions when the agreement offers the worker (employee) better conditions of employment (i.e. favourability principle).[203] It even goes as far as allowing for collective agreements to be extended to other limitations on certain constitutionally guaranteed rights. For instance, section 64(1)(a) prohibits strike where a collective agreement determines that the issue in dispute should not be subject to strike actions. Furthermore, by a collective agreement between an employer and a majority union, such a limitation may also be extended to workers who do not belong to the union concerned, thereby also depriving them of the rights to strike over that particular issue.

V. Conclusion and Recommendation

There is no gainsaying the fact that collective bargaining is a rational process for the enhancement of workplace democracy, redistribution of power from employers to employee, a forum for ascertaining and reviewing the terms and conditions of employment, and a veritable tool for the promotion of economic efficiency by limiting industrial conflict in the workplace. In general, consensus is that collective bargaining must be the nucleus of any dynamic modern system of industrial relations.[204] Notwithstanding, from the appraisal provided specifically dealing with collective bargaining, one can readily determine the level of protection that is accorded to the parties thereto.

As revealed in this study, the ILO has established core labour standards which enshrine workers' right to free and voluntary collective bargaining. Unfortunately, by global standards, collective bargaining practice appears to be in a dire state in the Nigerian labour sphere. The reasons seem not to be far-fetched. All over the world, the practice of industrial relations and collective bargaining emanated from the private sector. In Nigeria, the reverse is the case.[205] The reality is that the government has continued to pay lip service to mechanism of collective bargaining.[206] Whilst Nigeria has ratified the ILO Conventions, many of its practices concerning collective bargaining do not meet the ILO standards. For instance, the vagueness about the requirement of a certain level of

[201] Labour Relations Act, s 23 (1).
[202] ibid s 23 (3).
[203] Basic Conditions of Employment Act, 75 of 1996, s 49.
[204] Okene, 'The Challenges of Collective Bargaining in Nigeria' (n 1) 102.
[205] Olulu and Udeorah (n 5) 65.
[206] ibid.

representativeness in the form of an "electoral college" poses as a challenge to the recognition of workers' organisations and their inherent right to negotiate; certain class of public officers are not covered by collective bargaining; the scope of negotiable issues and the subject matter for collective bargaining are unjustly confined; the level of collective bargaining appears fictitious and is constantly plagued specifically in the public sector with administrative intrusions; there is the preponderance of interventionist policies and legislative attitude of compulsion and collective agreement seems readily unenforceable. The implication of these incongruences manifests in the form of deadlock collective bargaining process which continues to eat deep into the fabrics of Nigeria's labour sector with the frequent side-lining of the process by recourse to strikes and lock-out by the organised labour and management respectively. In sum, Nigeria's labour legal regime constricts and does not allow for the practice of collective bargaining to flourish. As long as these impediments highlighted above subsist, one can only in futility hope for a better inclusive collective labour legal landscape in Nigeria.

It is submitted that Nigeria must therefore make deliberate efforts in progressively overturning the hurdles on its way to achieving an internationalised labour regulatory framework. In a bid to authenticate the right to freedom of association and utilize its machinery of collective bargaining, the Nigerian government must therefore amend its laws to readily capture the standard prescriptions of the ILO as it relates to the practice of collective bargaining. Put specifically, the legislative focus should be geared towards enacting limited intervention guided democratic policies to reflect its commitment to voluntarism, workplace democracy, industrial peace and harmony.

Cambridge Law Review (2022) Volume VII, Issue 1, 155–167

Distributive Justice as a Function of Contract Law

JOHN (CHING JACK) CHOI*

ABSTRACT

According to the minimalist view of the function of contract law, contract law rules should merely facilitate the intentions of the bargaining parties, with a mind to respecting their autonomy. Considerations of distributive justice, accordingly, have no place in such a framework. This article argues that this view is incorrect. Contract law can and should have a more ambitious aim of promoting distributive justice by ensuring that private transactions achieve a fair distribution of wealth. Not only are distributional considerations already deeply embedded in contract rules, but contract law has the capacity and ability to perform a more robust distributive function.

Keywords: contract law; distributive justice; private law; legal theory; jurisprudence

I. INTRODUCTION

According to the conventional view, the function of contract law is merely to facilitate the intentions of the bargaining parties, with a mind to respecting their autonomy. Considerations of distributive justice, accordingly, have no place in such a framework.[1]

In disagreement with this view, this article argues that contract law can and should have a broader, more ambitious aim: to promote distributive justice by ensuring that private transactions achieve a fair distribution of wealth.[2] Section II

* Bachelor of Civil Law (BCL) candidate at the University of Oxford; LLB (UCL). With special thanks to Dr Lucinda Miller. All errors remain my own.

1 Aditi Bagchi, 'Distributive Justice and Contract' in Gregory Klass, George Letsas, and Prince Saprai (eds), *Philosophical Foundations of Contract Law* (Oxford University Press 2014) 195: "distributive justice is still perceived as not just misguided but *alien* to contract".

2 This article therefore assumes that contract law pursues *instrumental* goals. For a convincing rejection of anti-instrumentalist approaches to contract law, see Jonathan Morgan, *Contract Law*

will critically analyse those accounts proposed by advocates for a minimalist, facilitative function of contract law, and critique their underlying motivations. Building on that, Section III will sketch out the case that in a liberal society, contract law has the capacity and ability to perform such a distributive function. Finally, Section IV will investigate whether the adoption of a distributive lens makes any particular demands of the legal framework in which contract law's purpose is to be instrumentalised.

There is both a descriptive and normative aspect to the present inquiry. This article is *descriptive* to the extent that it proves how distributional considerations are already deeply embedded in contract law rules and therefore we cannot divorce them from the corpus of contract law even if we wanted to, as some scholars do.[3] It is *normative* to the extent that it requires contract rule makers, such as judges, to rethink their role and consider more readily and transparently the implications their decisions can have on the distributive arrangements between members of society.

II. Contract Law, Efficiency, and Wealth Maximisation

The idea of a purely facilitative function of contract law is inextricably linked to the theory of welfare maximisation and the value-neutral understanding of private law. According to the traditional dichotomy between the public realm (the state) and the private realm (the market), the state is there to support private ordering by supplying parties with legal enforcement to their private arrangements. The purpose of the institution of contract law is not to steer private ordering in any distributionally fair manner,[4] but to merely facilitate the desires of the parties.[5] Therefore, under the view that broadly coincides with most liberal theories, contract law's facilitative aim is constrained to the establishment and limited protection of the market. Any distributive consequences depend on the autonomous choices of the individual and, where appropriate, state intervention through taxation and the welfare system.

The central assumption underpinning this view is that private exchanges should be supported by contract law because people make themselves better off by pursuing their own self-interest. And, by extension, they make society better off. If efficiency is about the creation of a pie and distribution its division, then

Minimalism: A Formalist Restatement of Commercial Contract Law (Cambridge University Press 2013) chs 1 and 2.

3 Charles Fried, *Contract as Promise: A Theory of Contractual Obligation* (Harvard University Press 1981).

4 Ernest J Weinrib, *The Idea of Private Law* (Harvard University Press 1995) 80.

5 For a discussion of the conventional view that contract law only has a facilitative role: Hugh Collins, 'Regulating Contract Law' in Christine Parker and others (eds), *Regulating Law* (Oxford University Press 2004) 17.

according to this view, contract law should only be concerned with maximising the size of the pie, whereas taxation is tasked with its fair distribution amongst the guests at the dinner table.[6] To this end, there have been calls, most prominently made by scholars such as Jonathan Morgan, Alan Schwartz, and Robert Scott, for a formalist regime of contract law rules, one that is rid of judicial imagination and should do nothing more than facilitating parties' preferences.[7]

At the outset, there are at least two preliminary responses to be made against this simplistic view. Firstly, contract law rules, no matter how seemingly neutral, are fundamentally distributive in nature, and the assertion that a rule's welfare-maximising effects can be separated from its distributive implications is highly misleading. I will develop this point in Section III.A below.

Secondly, implicit in these formalist accounts is the distinction between an allegedly neutral private law arena (the market) and a value-laden, political realm (the state). As already noted, the conventional view holds that the state should only *facilitate* contracting in the marketplace. To go beyond that is to disrespect private autonomy. Yet, we may observe that this suggestion of a formalist contract law regime is reminiscent of a particular understanding of private law, one which was rejected by the American legal realist movement about a century ago.[8] When advocates speak of a politically neutral regime of contract law, they overlook the fact that courts are necessarily making policy choices about which contracts should or should not be enforced.[9] This is because typically when a party calls on the courts to uphold their agreement, it is inviting the state to coerce the other party to make good on her side of the bargain. As Morris Cohen influentially recognised,

> "[I]n enforcing contracts, the government does not merely allow the two individuals to do what they have found pleasant in their eyes. Enforcement, in fact, puts the machinery of the law in the service of one party against the other. When that is worthwhile and how that should be done are important questions of public policy."[10]

If we find that private contracting generates externalities that harm some segments of society more severely than others, then it is not only the (private) market players who are the culprits; rather, the public institution of contract law has also made a conscious policy decision to aid and abet that outcome. In other

[6] Louis Kaplow and Steven Shavell, *Fairness versus Welfare* (Harvard University Press 2002).

[7] See Morgan (n 2); Alan Schwartz and Robert E Scott, 'Contract Theory and the Limits of Contract Law' (2003) 113 Yale Law Journal 541.

[8] Morris Cohen, 'The Basis of Contract' (1933) 46 Harvard Law Review 553. See also Peer Zumbansen, 'The Law of Society: Governance Through Contract' (2007) 14 Indiana Journal of Global Legal Studies 191.

[9] Gary Peller, 'The Classical Theory of Law' (1998) 73 Cornell Law Review 300, 301–03.

[10] Cohen (n 8) 562.

words, given that players who have found themselves in a socially advantaged position can have their advantage reinforced and strengthened through the machinery of contract law, contract law cannot be neutral. Therefore, if those who are prejudiced under the current function of contract law can do better under a more distributively concerned regime, then such a change in perspective is very much commendable.

III. Contract Law as Exercising a Distributive Function

As the foregoing discussion shows, the enforcement of contracts is a deeply public and political phenomenon. Given that contract law is inseparable from the wider institutional arrangement, this section develops the idea that contract law can be used to address other normative concerns, beyond the adherence to freely chosen obligations: namely, to promote distributive justice.

Unlike corrective justice,[11] distributive justice takes a broader stance and is concerned with the allocation of wealth, resources, and entitlements amongst the members of any given society. Any attempt to usurp a distributive role from contract law invites several opposing arguments: that it is contradictory to the bilateral nature of private law; that it infringes on private autonomy; that it is futile and counterproductive; and that it is a less appropriate means compared to taxation. These arguments will be considered, and rejected, in turn in Sections II.B–II.E.[12]

A. Distributive justice throughout contract law

There is a "claim for neutrality" within contract law.[13] Consistent with the pre-political understanding of private (contract) law, this idea suggests that contract law rules merely enable private ordering and allow individuals to pursue

[11] Jules Coleman and Arthur Ripstein, 'Mischief and Misfortune' (1995) 41 McGill Law Journal 91, 93: "Corrective justice concerns the rectification of losses owing to private wrongs. In contrast, distributive justice concerns the general allocation of resources, benefits, opportunities, and the like."

[12] There is indeed more than one theory of distributive justice, and to pick one most suitable for contract law to adopt is beyond the scope of this article. The focus here is whether the machinery of contract law has good reason and the capacity to accommodate at least some theories of distributive justice.

[13] Hugh Collins, 'Distributive Justice Through Contracts' (1992) 45 Current Legal Problems 49, 49–52.

their own preferred goals; it does not impose any instrumental goals onto the parties.

But even if we could accept that contract law is this apolitical, value-neutral system of rules, separate from the upheavals of politics (a view which I have argued in Section II to be incorrect), it is impossible to ignore the distributional consequences that they entail. This was recognised by Anthony Kronman, who influentially observed that the laws of contract can have an important role to play in achieving a socially just allocation of resources, at least in some cases.[14] He gives the examples of usury laws, minimum wage laws, and so on.[15] I would go further and submit that *every* contract law rule, no matter how seemingly mundane and facilitative, possesses a distributive nature. In other words, there is no such thing as a non-distributional rule of contract.[16]

Consider the default rules in contract law. Although it is intuitive to attribute only *mandatory* rules with a regulatory dimension, as they are indeed the more intrusive and interventionist forms of law, *default* rules can also have strong implications for distribution, albeit less obviously. Even if parties contract out of default rules, as they often do, the rules still provide the parties a starting point that will determine *ex ante* how much each party will be able to bargain and which party will have to bear the main cost in deviating from the default structures. For example, in an incomplete contract where the parties failed to specify the consequences of delivering services that were below par than expected, there are at least two options open to the court. It may either: (a) inject a default rule to compel the service-provider to guarantee its quality; or (b) let the loss lie where it falls. A purely facilitative idea of contract law would most likely prefer the latter policy: after all, why 'regulate' the parties' relationship when they could have done so themselves? However, whilst this latter option resembles non-regulation, both options are, in fact, two different forms of regulation, each entailing its own distinct distributive arrangement.[17] On the one hand, a policy which protects the purchaser with a warranty bestows upon the purchaser a form of *ex ante* advantage even prior to the bargaining process. If the service-seller wants to substitute this implied warranty with her own preferred terms, she must bear the cost and give something in return, usually in the form of a lower price.[18] On the other hand, if the courts adopted a stance that lets the loss lie where it falls, that would give the service-seller an important benefit vis-à-vis the buyer. Similarly, if the buyer

[14] Anthony T Kronman, 'Contract Law and Distributive Justice' (1980) 89 Yale Law Journal 472.
[15] ibid 473.
[16] Marco Jimenez, 'Distributive Justice and Contract Law: A Hohfeldian Analysis' (2017) 43 Florida State University Law Review 1265, 1271 ("the idea of distributionally 'neutral' contract rules is a 'legal unicorn'"). Cf Fried (n 3).
[17] ibid 1308–11.
[18] ibid 1310.

wanted to insert such a warranty into the terms of the agreement, it is he who must do the convincing and shoulder the cost of bargaining.

The upshot here is that whenever the courts are asked to interpret an ambiguous term or make a decision in a contractual dispute, it essentially makes a distributive decision. In Marco Jimenez's words, "[e]very choice governing every rule in contract rule [*sic*] is a distributive choice setting the regime of background rules against which the parties bargain with one another".[19] Of course, one might question whether the fact that contract rules have a distributive *effect* should necessarily mean that one of the *functions* of contract law should be about fair distribution.[20] But, as Hugh Collins questions, once we accept that contract law has foreseeable distributive effects on the market, "[h]ow can it then be maintained that these foreseeable effects are not intended effects, that is, effects which are not part of the purpose of the law of contract?"[21] It is therefore not unmeaningful to ask the question of whether contract law is indeed suitable to pursue this distributive purpose, and if so, how its encompassing rules should go about generating distributive effects in an equitable manner.

B. The 'problem' of bilateralism in private (contract) law

One argument against the use of contract law for distributive purposes is that it would run counter to the very nature of private law. The fact that bilateralism is such a defining feature of private law means that contract law struggles to take into account the interests of parties outside the contract.[22] As there is no obvious link between the collective values imbued in distributive justice and the bilateral contracting parties, private law seems therefore a poor tool for redistribution. Yet, this argument rests on a fallacy that contracting parties are somehow insular to the community as a whole. We are reminded by Aditi Bagchi that "[c]ontracting parties do not encounter each other in a legal or moral vacuum" and that the "morality of exchange, agreement and even promise… are contingent on institutional arrangements".[23] The contractual relationship between parties, whilst mostly private and self-invented, are nonetheless informed by other

[19] ibid 1306. See also Collins (n 13) 65–67.
[20] Jane Stapleton, 'Regulating Torts' in Parker and others (n 5) 124.
[21] Collins (n 13) 51.
[22] Aditi Bagchi, 'Other People's Contracts' (2015) 32 Yale Journal on Regulation 211.
[23] Bagchi (n 1) 199 (emphasis omitted).

interpersonal duties which are in turn informed by the state of distribution in the outside world.[24]

Nonetheless, there is still one way in which it might be more accurate to say that bilateralism poses a problem: that is, contract law rules do not actively consider the interests of other parties. This is the externality argument. At least two reasons, however, can be offered in defence of contract law, each drawn to some extent from the increased blurring of the public-private divide.[25] Firstly, courts have demonstrated an ability to consider the collective interests of a group or class of people who share characteristics similar to that of the parties before it, and this informs their legal reasoning.[26] This is evident, for example, in the differing approaches of the Court of Appeal and of the House of Lords in *Co-operative Insurance Society Ltd v Argyll Stores (Holdings) Ltd*.[27] When considering whether to issue an injunction to maintain the opening of a supermarket in a shopping mall until other tenants are found, the former court took the view that it is the tenants of the shopping mall whose interests need protecting. By contrast, the latter held that it is the supermarket tenants who should prioritised. The precise reasoning of each proposition is less relevant for our purposes, but the important point here is that contract law judges can and do have the capacity to assess the distributive impact their decisions can have beyond the bilateral contracting parties.[28]

Secondly, though it is conventional to think of third-party effects as only relevant at the *ex ante* legislation stage, rather than in the *ex post* adjudication stage, there is nonetheless a sense in which the public 'participates' in private adjudication.[29] Through interpretation[30] and utilising considerations of public policy, the courts, at least in some cases, can manage, adjust, and recalibrate the social ramifications caused by the contractual obligations of private parties.

C. Interference with Party Autonomy

A second, yet related, objection to the use of contract law for distributive purposes is that such a policy entails an undue interference with the principle of freedom of contract, which Lon Fuller once observed to be "the most pervasive

[24] ibid.

[25] On the public-private distinction, see generally Duncan Kennedy, 'The Stages of the Decline of the Public/Private Distinction' (1982) 130 University of Pennsylvania Law Review 1349.

[26] Hugh Collins, *Regulating Contracts* (Oxford University Press 1999) 70–73.

[27] *Co-operative Insurance Society Ltd v Argyll Stores (Holdings) Ltd* [1996] Ch 286 (CA); [1998] AC 1 (HL).

[28] For other judicial and academic examples of a class-based analysis, see *Smith v Eric S Bush* [1990] 1 AC 381 (HL); Orit Gan, 'The Justice Element of Promissory Estoppel' (2015) 89 St John's Law Review 55, 86–87.

[29] David A Hoffman and Cathy Hwang, 'The Social Cost of Contract' (2021) 121 Columbia Law Review 979, 985.

[30] ibid 1002–05.

and indispensable" conception of contract law.[31] The argument, which broadly coincides with the libertarian view, contends that contract law should not, even in the name of promoting fair distribution, interfere with transactions freely consented to by private parties.[32] A closer examination of contractual freedom, however, will show that it may pose less of an issue than first thought.

In the first place, Peer Zumbansen's contention that "[t]here was never a period of pure freedom of contract or of pure private autonomy" is telling.[33] Contrary to classical theory, private contracting has necessarily evolved in the context of some regulatory framework, constrained to varying extents by notions of fairness and justice. The alleged misalignment between the pursuance of distributive justice on the one hand, and the respect for party autonomy on the other, is reduced when we consider that the two are invariably interdependent.[34] In reconciling the concepts of autonomy and fairness, Florian Rödl said, "[t]here is no tension between the two concepts of contractual freedom and contractual justice because contractual freedom can only be exercised in voluntary agreements with fair terms. Unfair contracts cannot be claimed valid by appealing to contractual freedom".[35]

In the second place, not all legal doctrines that restrict contractual freedom fit equally and neatly in the notion of 'a search for real consent'. Contract law tends to marginalise the extent to which the courts can and do assess the substantive fairness of bargains by couching freedom-restraining rules in the language of realising parties' true intentions. Yet, amongst these doctrines, there are several that can be explained in parallel with distributive language, and then there are others where adopting a pure non-distributive explanation would be quite unpersuasive. Rules of procedural fairness such as the doctrine of misrepresentation are examples of the former.[36] Whilst it is entirely appropriate to rationalise misrepresentation as vindicating the autonomy of the promisee who was misled into entering an agreement she did not truly consent to, it can also be conceived as a tactic for redistributing wealth in the market.[37] The rule restrains dominant parties from exploiting the market advantages they have, whether it be superior product information or intellect, to the detriment of others without the opportunity to gain such advantages.

31 Lon L Fuller, 'Consideration and Form' (1941) 41 Columbia Law Review 799, 806.

32 *Printing and Numerical Registering Co v Sampson* (1875) 19 LR Eq 462 (CA) 465.

33 Zumbansen (n 8) 207. See also P S Atiyah, *An Introduction to the Law of Contract* (5th edn, Clarendon 1995) ch 1.

34 Florian Rödl, 'Contractual Freedom, Contractual Justice, and Contract Law (Theory)' (2013) 76 Law and Contemporary Problems 57, 62.

35 ibid. Although Rödl does not speak specifically of 'distributive justice', much of his discussion on 'contractual justice' aligns with the substance of the former concept.

36 Collins (n 26) 75.

37 Kronman, (n 14) 480–82.

As to the latter group of rules, the rules of contractual interpretation is a likely contender. The notion of freedom of contract provides only partial guidance as to how courts decide what is the best contractual reading of a concluded agreement. Although Sarah Worthington warns the courts against implying or reconstructing terms too readily on the grounds of substantive unfairness,[38] it is clear that when tasked with construing contracts, the courts are not just engaged in an exercise of reassembling bargains as they once were.[39] Bagchi's quadripartite account of what courts consider 'reasonable' (empirically, procedurally, substantively, and publicly) reminds us that the search for the best contractual reading is motivated also by a desire to ensure transactions achieve a just allocation of wealth.[40]

D. Futility and Counterproductivity

A third set of objections contends that the pursuance of distributive goals through contract law can be futile and may even be counterproductive.

In terms of futility, proponents of a facilitative, minimalist function of contract law argue that the attainment of distributional goals through the law would not produce meaningful results. Schwartz and Scott contend that the reason for this is because most parties will exercise their freedom to contract away distributive rules that do not serve their goals of profit maximisation.[41] This is not persuasive. Firstly, as Schwartz and Scott themselves admit, the force of such an argument greatly diminishes once we leave the realm of commercial (firm-to-firm) contracts. Although commercial contracts arguably form the bulk of headline contractual disputes, they are but a subset of everyday contracting. Their narrow account ignores, for instance, consumer contracts, employment contracts, or government contracts[42], and therefore cannot be generalised. In these other types of contracts, distributive concerns come even more to the forefront, thereby making it less likely and less desirable for parties to contract out of distributively-motivated terms. But secondly, even if we do stick *within* the realm of commercial contracts, that most contract rules are easily overridden does not mean that any

[38] Sarah Worthington, 'Common Law Values: The Role of Party Autonomy in Private Law' in Andrew Robertson and Michael Tilbury (eds), *The Common Law of Obligations: Divergence and Unity* (Hart Publishing 2016) 319.

[39] S M Waddams, 'Unconscionability in Contracts' (1976) 39 MLR 369, 382–84. See also *L Schuler AG v Wickman Machine Tool Sales Ltd* [1974] AC 235 (HL); *Yam Seng Pte Ltd v International Trade Corp Ltd* [2013] EWHC 111 (QB), [2013] 1 All ER (Comm) 1321.

[40] Aditi Bagchi, 'Interpreting Contracts in a Regulatory State' (2019) 54 University of San Francisco Law Review 35, 73.

[41] Schwartz and Scott (n 7) 545–46.

[42] On the inadequacy of efficiency theory in the context of government contracting, see Wendy Netter Epstein, 'Contract Theory and the Failures of Public-Private Contracting' (2013) 34 Cardozo Law Review 2211.

pursuance of distributive objectives must therefore be futile. As discussed in Section III.A above, even when firms sidestep default rules, they nonetheless provide an important *ex ante* starting point for the bargaining process to take place.

Another argument on similar lines is that redistribution through the regulation of contracts may even be counterproductive to the goals it purports to achieve. Namely, it risks raising transaction costs at the detriment of the parties who are the intended beneficiaries of the distributive policy.[43] But although it may well be the case that costs are sometimes passed on, this is not a reason in itself for contract law to cease offering such protections altogether. Indeed, "[r]ules intended to advance distributive justice exact a cost".[44] To use the above example of a service-seller and buyer,[45] if the imposition of a pro-buyer default rule makes it more costly for the buyer to enter into the contract, she may nonetheless consider it worthwhile to do so for the benefit of an added layer of legal protection. That being said, it would of course be most unacceptable if that cost was borne disproportionately by an already socially disadvantaged group. Yet, whether this occurs is a highly fact-sensitive question, and the issue requires careful consideration by rule-makers.[46] It should not support a blanket aversion to redistributive restrictions on contractual freedom, especially in the light of evidence exposing the exaggerated nature of some regulatory-backfiring arguments.[47]

E. Taxation: A Better Alternative for Redistribution?

Finally, even if we accept that the machinery of contract law has the capacity to incorporate distributive considerations within its framework and good reasons to do so, it may nonetheless be open for advocates for distribution to hold that, given the relative advantages of the tax and welfare system, the role of wealth redistribution should ultimately be kept away from contract law.[48] However, we should be careful not to lose sight of the weakness inherent in public techniques for redistribution. At a theoretical level, the contention that taxation is less intrusive or more morally justifiable should be subjected to scrutiny;[49] and at a practical level, the assertion that taxation is inevitably more efficient can be

[43] Morgan (n 2) 153–156.
[44] Bagchi (n 1) 210.
[45] See Section III.A.
[46] Bagchi (n 1) 210.
[47] Collins (n 26) 277–79. Cf Morgan (n 2).
[48] Advocates for this view include Michael Trebilcock. See Michael J Trebilcock, *The Limits of Freedom of Contract* (Harvard University Press 1993).
[49] For an objection to the proposition that redistribution through taxation is more morally permissible than contract law, see Kronman (n 14) 498–507.

challenged. As Kronman observes, there is always at least one way in which the law of contract is more efficient than taxation: unlike taxation, which requires the prior collection of state revenue before redistribution, contract law, through its rules and doctrines, can facilitate a direct transfer of resources and advantages from one part of the community to another without additional state mediation.[50]

IV. Towards a Law of Contract More Apt to Promote Distributive Justice

If contract law were to adopt a more robust, distributive function, how should its doctrinal framework look like to instrumentalise that goal? A detailed re-examination of the doctrines of contract law is beyond the scope of this article; however, this last section seeks to put forward a general direction in which the law could follow to promote distributive justice in the contracting process. If one overarching theme had to be identified, it would be that contract law doctrines should pay more attention to the structural inequalities between the parties, the allocative impacts of substantive terms, and other public policy considerations. For reasons discussed below, this has particularly significant implications for the empowerment of traditionally disadvantaged social groups, who will overall benefit from a more inclusive, equitable, and pluralist law of contract.[51] Three principles are suggested here.

Firstly, a distributive contract law should be sensitive to how its rules have different implications for different segments of the population. Consider the requirement for intention to create legal relations, which has been the subject of scathing feminist critiques.[52] It has been revealed that this doctrine, and its presumption against legal bindingness in the domestic setting, has the effect of "[insulating] the female world from the legal order", as it "devalues women by saying that they are not important enough to merit legal regulation".[53] Given that doctrines such as this one often deal with members of different social groups, contract law should take into account the power imbalances and allocative consequences of the contracting process. For instance, contract law could do away with this presumption of legal bindingness altogether or lower the threshold for rebutting the presumption. The end result would be to enable the courts to steer private ordering in a more distributively equitable fashion and empower the underprivileged by giving them greater bargaining power. A refusal to do so

[50] ibid 509.

[51] See Patricia J Williams, *The Alchemy of Race and Rights* (Harvard University Press 1991).

[52] See Michael Freeman, 'Contracting in the Haven: *Balfour v Balfour* Revisited' in Roger Halson (ed), *Exploring the Boundaries of Contract* (Dartmouth 1996).

[53] ibid 74.

would have detrimental distributive ramifications, as it would reinforce the status quo and preserve the economic and social inequalities between any two social groups, in this case by privileging men over women.

Secondly, when interpreting or construing ambiguous agreements, the courts should include distributive considerations as relevant factors. A literal, four-corners approach ought to be resisted as it is too insensitive to the allocative implications of everyday contracting. In a sense, this recommendation is hardly novel, as both scholars[54] and modern-day courts have advocated for a contextual approach to interpretation. However, even under most contextual approaches, the scope for promoting distributively fair outcomes is limited, as the focus is still on unearthing the parties' true intentions. Contextualists only differ from textualists in terms of methodology, the former being more liberal in respect of relying on extrinsic evidence than the latter.[55] Instead, the courts should be willing to prefer some contractual readings over others on the explicit ground that, for example, it could lead to more equitable distributions or fairer impacts to third parties.[56]

Thirdly, contract law should do away with its traditional antipathy towards regulating substantive fairness.[57] In congruence with what has been said above and in line with developments in other common law jurisdictions,[58] it is submitted that English law should recognise a doctrine of unconscionability as a general ground for relieving parties of certain perverse contractual obligations. This is no invention: the courts are already doing it under the guise of other rules, such as the rules on the incorporation of terms, duress, and the regulation of exemption clauses.[59] Open recognition can allow the courts to intervene more readily in bargains where one party seeks to exploit a distributive injustice, as well as to make their decisions more transparent. As Stephen Waddams asserts, "despite lip service to the notion of freedom of contract, relief is every day given against agreements that are unfair, inequitable, unreasonable or oppressive".[60] This principle of unconscionability could be invoked, for instance, where the court finds that there is substantive unfairness in the terms of the contract. The effect of such a finding would be to vitiate the contract in the same way as misrepresentation does, thereby granting the complainant the option to set aside the contract.

54 Particularly relational scholars. See Hugh Beale, 'Relational Values in English Contract Law' in David Campbell, Linda Mulcahy, and Sally Wheeler (eds), *Changing Concepts of Contract: Essays in Honour of Ian Macneil* (Palgrave Macmillan 2013).

55 Bagchi (n 40) 37–38.

56 Bagchi (n 1) 208.

57 Collins (n 13) 65–67.

58 More prominently in American jurisdictions.

59 Waddams (n 39).

60 ibid 390.

V. Conclusion

The question that all contract rule-makers should ask themselves is this: "what kind of contract law should the state offer?" A facilitative regime of contract law might have the benefit of consistency, clarity, and predictability, and is perhaps more easily applied than a context-sensitive one. Yet, in our factually varied world, "the moral reality of distributive injustice"[61] requires any meaningful conception of contract law to not only blindly facilitate the exchanges between private parties, but also to consider broader considerations, such as their power dynamics, background duties, as well as allocative implications.

[61] Bagchi (n 1) 210.

Cambridge Law Review (2022) Volume VII, Issue 1, 168–180

The Retained EU Jurisdiction to Suspend Remedies in English and Welsh Law: *R (Open Rights Group and the3million) v Secretary of State for the Home Department and Secretary of State for Digital, Culture, Media and Sport* [2021] EWCA Civ 1573

SAMUEL WILLIS*

ABSTRACT

Before the United Kingdom withdrew from the European Union (EU), domestic courts did not have the discretion to suspend public law remedies. Such a discretion did exist within the sphere of EU law, the exercise of which lay with the Court of Justice of the European Union (CJEU). The European Union (Withdrawal) Act 2018 purported to translate directly applicable EU law, with specified omissions, into domestic law (retained EU law). As a result, for the first time, UK courts acquired the discretion to suspend the effect of public law remedies, such as quashing orders, albeit within the sphere of retained EU law.

In *R (Open Rights Group and the3million) v Secretary of State for the Home Department and Others* [2021] EWCA Civ 1573 (*Open Rights Group (No 2)*), the Court of Appeal faced an application to exercise this discretion to suspend the disapplication of the "Immigration Exemption" in the Data Protection Act 2018. In answering this application, Lord Justice Warby gave shape to this new domestic jurisdiction. The judgment is significant for three key reasons. Firstly, it identifies an anterior question to be answered when English and Welsh courts are called upon to enforce a rule of retained EU law—namely, whether the rule of law in

* Lord Denning Scholar, Lincoln's Inn, and student at the Inns of Court College of Advocacy. BA (Cantab), M.Phil (Cantab), and GDL (City Lond). I am grateful to Professor Ian Loveland (City, University of London) for bringing this case to my attention. Any errors that remain are my own. samuelawillis@aol.com.

issue was capable of translation into English and Welsh law. Secondly, Warby LJ's judgment appears to model, if not explicitly identify, the correct approach to answering that question. Finally, the judgment provides guidance for the exercise of the discretion. Despite claiming to adhere to the CJEU's high threshold for exercising the jurisdiction, close analysis of the Court's reasoning appears to indicate a lowered threshold for its exercise by English and Welsh courts in the sphere of retained EU law.

Keywords: retained EU law; suspended remedies; Brexit; European Union (Withdrawal) Act 2018; legal certainty

I. Introduction

In 2008, in *Kadi v Council and Commission*, the European Court of Justice (ECJ) annulled sanctions imposed on Mr Kadi, who was suspected of having funded al-Qaeda.[1] The annulment, however, was to be suspended for "a brief period" to "allow the Council to remedy the infringement found".[2] The Court took this step because it considered that immediate annulment "would be capable of seriously and irreversibly prejudicing the effectiveness" of the sanctions regime.[3] There existed no counterpart to this jurisdiction in domestic English and Welsh law. This position changed with the enactment of the European Union (Withdrawal) Act 2018 (EUWA 2018), which "photocopied" the corpus of European Union (EU) law on the Implementation Period completion day (IP completion day) and purported to translate it with specified omissions into domestic law.[4] As such, for the first time, domestic United Kingdom (UK) courts acquired a domestic statutory jurisdiction to suspend relief in public law challenges, albeit within the sphere of retained EU law.

It is perhaps ironic that the UK Parliament, not long after the UK's withdrawal from the EU, is considering a proposal to create a general jurisdiction to suspend public law remedies. Clause 1 of the Judicial Review and Courts Bill would empower UK courts to suspend quashing orders (cl. 1(1)(a)) or remove or limit any retrospective effect of the quashing (cl. 1(1)(b)). This is the context in which this judgment arrives; it seems that, despite Brexit, suspended remedies will be a feature of UK public law for the foreseeable future. Lord Anderson of Ipswich,

1 Case C-402/05, *Kadi and Al Barakaat International Foundation v Council and Commission* [2008] ECR I-6351, ECLI:EU:C:2008:461, [373]–[376].

2 ibid [375]. The suspension was to last for three months, see ibid [376].

3 ibid [373].

4 For the IP completion day, see: the European Union (Withdrawal Agreement) Act 2020, s.39(1), and the European Union (Withdrawal) Act 2018, s.1A(6). For the translation of EU law into domestic law as "retained EU law', see: the European Union (Withdrawal) Act 2018, ss 2-7.

a veteran advocate in the Court of Justice of the European Union (CJEU), has welcomed this: "[p]erhaps because I have become used to these remedies in practice, I believe that each has its place, if not at the top of the judicial toolbox, then certainly somewhere within it".[5]

In this emerging area of English and Welsh law, *R (Open Rights Group and the3million) v Secretary of State for the Home Department and Secretary of State for Digital, Culture, Media and Sport* [2021] EWCA Civ 1573 (*Open Rights Group (No 2)*) stands as an important intervention by the Court of Appeal.[6]. Since the case is relatively recent, it is unsurprising that it has not yet been the subject of extensive academic study. It may also be that *Open Rights Group (No 2)* has been overshadowed by *R (Open Rights Group and the3million) v Secretary of State for the Home Department and Others* [2021] EWCA Civ 800 (*Open Rights Group (No 1)*), in which the Court of Appeal identified the incompatibility that gave rise to *Open Rights Group (No 2)*. This article seeks to address this gap by teasing out the three key lines of reasoning in Warby LJ's judgment: (a) the identification of the anterior question of whether an EU rule of law is capable of translation into English and Welsh law; (b) Warby LJ's modelling of the approach to be taken in answering that question; and (c) the providing of guidance for English and Welsh courts in the exercise of the retained EU discretion to suspend public law remedies.

II. History of the Proceedings

The appellants brought the judicial review claim against the Home Secretary and the Secretary of State for Digital, Culture, Media and Sport in August 2018. The appellants were two non-governmental organisations (NGOs): the Open Rights Group, a digital rights NGO; and the3million, a grassroots organisation representing EU citizens resident in the UK. The two organisations sought a declaration that the "Immigration Exemption" in paragraph 4 of Schedule 2 to the Data Protection Act 2018 (DPA 2018) was non-compliant with Article 23 of the UK General Data Protection Regulation (GDPR). The application was first heard by Supperstone J, who dismissed the application in October 2019.[7] Singh LJ

[5] HL Deb 7 February 2022, vol 818, col 1351.

[6] *R (Open Rights Group and the3million) v Secretary of State for the Home Department and Secretary of State for Digital, Culture, Media and Sport* [2021] EWCA Civ 1573 (*Open Rights Groups (No 2)*).

[7] *R (Open Rights Group and the3million) v Secretary of State for the Home Department and Secretary of State for Digital, Culture, Media and Sport* [2019] EWHC 2562 (Admin).

granted leave to appeal in November 2019. The appeal was heard by Lord Justice Underhill V-P, and Lord Justices Singh and Warby.

The Immigration Exemption disapplies some data protection rights where the application of those rights would be likely to prejudice immigration control.[8] Article 23 of the GDPR authorises such exemptions. The status of the GDPR in domestic law is clear. The EU GDPR was translated directly into English law as the UK GDPR. It retains supremacy over other domestic instruments enacted before IP completion day.[9] This means that conflicts between the UK GDPR and other domestic legislation enacted prior to IP completion day, including primary legislation such as the DPA 2018, must be resolved in favour of the GDPR.

The issue, then, was whether the Immigration Exemption in the DPA 2018 is compatible with Article 23 of the GDPR. This provision authorises exemptions from certain data protection rights through a "legislative measure" where the exemption "respects the essence of fundamental rights and freedoms" to safeguard "specific objectives". These objectives are set out in Article 23(1)(a) to (j). Besides these specific objectives, Article 23(1)(e) also permits the safeguarding of "other important objectives of general public interest". At first instance, Supperstone J found that the exemption was a matter of "important public interest". For this reason, the Immigration Exemption in the DPA 2018 was found to be compliant with the GDPR. On appeal, the Court of Appeal unanimously held that the Immigration Exemption did not fall within the scope of authorised derogations in Article 23 GDPR.[10]

Having made this decision, the Court of Appeal acknowledged that there arose the issue of whether relief could, and if so should, be suspended. Arrangements were made for a separate hearing, which resulted in the judgment at issue in this note.

III. Jurisdiction to Suspend Relief

Where a court finds that national primary legislation is incompatible with retained EU law, the appropriate remedy is declaratory relief.[11] This was the case during the UK's membership of the EU. It is not, and was not, constitutionally possible for domestic courts to quash primary legislation. Instead, courts could make a declaration to the effect that the incompatible provision in the primary legislation

[8] *R (Open Rights Group and the3million) v Secretary of State for the Home Department and Secretary of State for Digital, Culture, Media and Sport* [2021] EWCA Civ 800 [1] (Warby LJ) (*Open Rights Group (No 1)*).

[9] European Union (Withdrawal) Act 2018, s.5(2).

[10] *Open Rights Group (No 1)*, ibid [53]–[54].

[11] *R v Secretary of State for Transport, ex p Factortame Ltd and Others* [1990] 2 AC 85; [1991] 1 AC 603; [1992] QB 680. See also *R (Miller) v Secretary of State for Exiting the European Union* [2017] UKSC 5; [2018] AC 61 [67].

had been disapplied.[12] This was the remedy sought by the appellants. It was not disputed that this disapplication declaration remedy subsisted after the UK's withdrawal from the EU with regard to retained EU law. At issue was whether the CJEU's jurisdiction to suspend disapplication of inconsistent domestic law had been translated into English and Welsh law.

The origins of this jurisdiction can be identified in domestic and EU case law. Lord Mance JSC had, *obiter*, observed that such a suspension would be possible in *R (Chester) v Secretary of State for Justice*.[13] In *R (National Council for Civil Liberties) v Secretary of State for the Home Department and Anor*, the Divisional Court found that such a jurisdiction existed and thought it appropriate to exercise it.[14] *Liberty* concerned the inconsistency of Part 4 of the Investigatory Powers Act 2016 with the e-Privacy Directive 2002/58. The inconsistency in Part 4 was twofold. First, it permitted the retaining of communications data in the area of criminal justice without limiting its use to combating "serious crime". Second, it permitted this retention without regard to the requirement in EU law of prior review by a court or independent administrative body. The common thread between *Liberty* and the instant case, *Open Rights Group*, is the failure of domestic law to comply with procedural safeguards found in (retained) EU law. It is the inconsistent absence of law, not the presence of inconsistent law, that required remedy. Disapplying an inconsistent provision is safely within the judicial function; inserting a safeguard scheme to achieve consistency is not.

Since *Liberty*, the Grand Chamber of the CJEU has recognised and given shape to this jurisdiction in three cases. These authorities are *La Quadrature*, *Gewestelijke*, and *B v Latvia*.[15] *La Quadrature* is the key EU authority for this suspensory jurisdiction. Its *ratio* is in two parts: (a) "where a subsidiary rule of (national) law is inconsistent with a dominant rule of (EU) law and must therefore be overridden, there must be a judicial power to delay the implementation of the dominant rule, where that is necessary for compelling reasons of legal certainty"; but (b) "in the interests of legal certainty, that judicial power must be reserved to the CJEU".[16] The first element of this jurisdiction establishes the power and the condition for its exercise—"for compelling reasons of legal certainty". This limb does not pose any issue of translation into English and Welsh law. The second limb does pose a problem. As Warby LJ notes, "slavishly literal application of the second

[12] *R v Secretary of State for Employment, ex p Equal Employment Commission* [1995] 1 AC 1.

[13] *R (Chester) v Secretary of State for Justice* [2013] UKSC 63, [2014] AC 271 [72]–[74] (*Chester*).

[14] *R (National Council for Civil Liberties) v Secretary of State for the Home Department and Anor* [2018] EWHC 975 (Admin), [2019] QB 481 [17] (*Liberty*).

[15] Cases C-511/18, C-512/18, and C-520/18, *La Quadrature du Net and Others* [2021] 1 CMLR 31; Case C-24/19, *A v Gewestelijke Stedenbouwkundige Ambtenaar van het Department ruimte Vlaanderen* [2021] CMLR 9; Case C-439/19, *B v Latvijas Republikas Saeima* [2022] 1 CMLR 9, ECLI:EU:C:2021:504.

[16] *Open Rights Group (No 2)* (n 6) [27]. See Cases C-511/18, C-512/18, and C-520/18, *La Quadrature du Net and others* [2021] 1 CMLR 31.

element of the ratio would defeat the first" element. Since UK courts can no longer make preliminary references to the CJEU, the literal translation of the second limb into domestic law would render the jurisdiction a nullity.[17] The Court was therefore required to answer a further question: if the jurisdiction was retained in domestic law, what form did it take?

IV. The Status of Retained EU Law

These two issues arise from the blanket retention of EU law, case law, and general principles of law in the EUWA 2018. The Court of Appeal considered the mechanism of retention "clear enough".[18] First, UK courts must now decide issues as to the validity, meaning or effect of retained EU law for themselves; they are no longer able to make references to the CJEU.[19] Second, the general rule is that UK courts are to decide any such questions in accordance with relevant retained case law and principles of EU law (EUWA, s.6(3)). "Retained case law" and "retained general principles" are those principles established and decisions made before IP completion day. UK courts are not bound but "may have regard" to those principles established or decisions made after IP completion day.[20]

This is the general position. A special set of rules, however, apply to a "relevant court", of which the Court of Appeal is one.[21] Subject to one of the exceptions (none of which applied in this case), "relevant courts" are not absolutely bound by retained EU case law.[22] The test for a relevant court to depart from EU retained law is "the same test as the Supreme Court would apply in deciding whether to depart from the case law of the Supreme Court".[23] Lord Gardiner LC laid down this test in *Practice Statement*: the Court may "depart from a previous decision when it appears right to do so", but it "will bear in mind the danger of disturbing retrospectively the basis on which contracts, settlements of property and fiscal arrangements have been entered into and also the special need for certainty as to the criminal law".[24] Since *La Quadrature* and *Gewestelijke* were decided before IP completion day, the Court of Appeal was not in the instant case absolutely bound by them but was required to decide the case in accordance with them unless

17 *Open Rights Group (No 2)*, ibid.
18 ibid [23].
19 EUWA 2018, s.6(1)(b).
20 EUWA 2018, s.6(1) and (2).
21 The European Union (Withdrawal) Act 2018 (Relevant Court) (Retained EU Case Law) Regulations 2020 (SI 2020/1525), Regulation 3(b) (Regulations 2020).
22 EUWA 2018, s.6(4)(ba); and the Regulations 2020, Regulations 1 and 4.
23 EUWA 2018, s.6(5A)(c); and the Regulations 2020, Regulation 5.
24 *Practice Statement (Judicial Precedent)* [1966] 1 WLR 1234.

it felt it was right to depart from them.[25] *B v Latvia* was decided after IP completion day. Accordingly, the Court of Appeal was able to "have regard" to it.[26]

V. The Form of The Retained Jurisdiction to Suspend Relief

This legislative scheme for the retention of EU law provides three possible approaches to this suspensory jurisdiction. The first is that, since only the CJEU could exercise the power in *La Quadrature* and *Gewestelijke* to suspend relief regarding substantive laws, the retained suspensory power cannot be exercised by UK courts. This would be based on a literal reading of the retained EU case law. Warby LJ rejects this as "unduly mechanistic" and "tending to subvert rather than promote the legal policy that underlies this aspect of the CJEU jurisprudence".[27] This reasoning has potential wider implications for the body of retained EU law. The approach suggested by Warby LJ is a purposive one of adapting retained EU law to give effect to the underlying legal policy of the rule of law in issue.

The second option is to exercise the Court's power as a "relevant court" under Regulation 5 to depart from the second limb of *La Quadrature*.[28] Warby LJ considers that, following Lord Gardiner LC's test, such a departure was permissible because it would not cause "legal disorder".[29] Warby LJ, however, prefers a third approach, which is conceptually anterior to the potential exercise of the power in Regulation 5. Warby LJ considers the second limb "simply incapable of direct transposition into the domestic legal order as it now stands".[30] According to this analysis, the power to depart from EU case law might be exercised in relation to the first limb of *La Quadrature*. This piece of EU case law was transposed into English and Welsh law. It cannot be exercised in relation to the second limb, however, since the second limb was not translated into domestic law. Regarding the second limb, there is no retained EU law from which to depart.

In this way, Warby LJ's judgment is not merely an authority for the exercise of this suspensory jurisdiction in domestic law; it is an authority for determining whether principles of EU law or case law are capable of translation into UK domestic law. This is conceptually prior to any question of whether to depart from retained authorities. This is significant because it means that such analysis may be deployed by courts other than the "relevant courts" in Regulation 3. The purpose

25 *Open Rights Group (No 2)* (n 6) [24].
26 ibid.
27 ibid [27].
28 The Regulations 2020 (n 19) Regulation 5.
29 *Open Rights Group (No 2)* (n 6) [28].
30 ibid.

of limiting the power to depart from retained EU case law was, presumably, to reduce scope for legal uncertainty. Depending upon the creativity of lawyers and judges in first-instance courts and tribunals, such certainty may now be dependent on the creation of a body of domestic case law on the translation of EU law.

The proper approach to this anterior analysis is unclear. Warby LJ notes that, in the instant case, both the "departure route" and "non-translation route" are possible.[31] For this reason, he felt it was "not necessary to reach a definitive conclusion on the matter".[32] More than this, he considered it "better not to do so".[33] This is a missed opportunity to give guidance to lower courts that do not possess the Court of Appeal's power to depart from retained EU case law. For now, first-instance judges and lawyers will need to infer the proper approach from Warby LJ's analysis. First, Warby LJ identifies the purpose of the rule of law in question, the second limb of *La Quadrature*: to "ensure the law is interpreted and applied uniformly across the Union".[34] Temporary suspension of relief was only possible with the approval of the CJEU. The second step is a comparative analysis of the EU and English and Welsh legal systems. Outside the EU, the underlying policy of the second limb falls away. As Warby LJ notes, the English and Welsh system does not rely on a system of referrals on points of law: "courts and tribunals at all levels are duty bound to decide legal issues on which there is no precedent that binds them".[35] In Warby LJ's view, this comparative structural analysis of the EU and English and Welsh systems leads to the conclusion that this principle "has not been translated because it cannot be translated".[36] Before IP completion day, first-instance courts and tribunals were able to suspend relief when so authorised by the CJEU via the preliminary reference procedure. The falling away of the second limb of *La Quadrature* leaves first-instance courts and tribunals free "in principle" to suspend relief.[37]

For first-instance courts, which otherwise do not have the power to depart from EU case law, it would seem that the proper approach to the issue of "non-translation" of an EU rule of law is: (a) to identify the purpose of said EU rule of law; (b) to situate it in the framework of the EU legal system; (c) to identify the means by which the same result is achieved in the English and Welsh system; and (d) to find that the EU rule of law has not been translated if it can be mapped onto an existing feature of the English and Welsh system. Obsolete, non-translated EU rules of law may include other fetters imposed by the CJEU on national courts to

[31] ibid [31].
[32] ibid.
[33] ibid.
[34] ibid [29].
[35] ibid.
[36] ibid.
[37] ibid.

protect the unity and uniformity of EU law. The second limb of *La Quadrature* is an example of such a fetter.

VI. The Exercise of The Retained Jurisdiction

Having established the retention of this jurisdiction in modified form, the Court considered the test for the exercise of this power. Linked to this test is the guidance for when to exercise this power. The following test can be derived from the retained authorities, *La Quadrature* and *Gewestelijke*: (a) this jurisdiction should be exercised exceptionally on the basis of "overriding considerations of legal certainty"; (b) the interests of legal certainty must be so compelling that it is necessary "for them to take priority over the need to implement the dominant legal provision, and disapply the subordinate law";[38] (c) this means that immediate disapplication would cause "serious difficulties" with respect to legal certainty; and (d) the party seeking to rely on the suspension has acted in good faith. The latter two propositions are derived from *B v Latvia*, which was decided after IP completion day. Nonetheless, the Court of Appeal "had regard" to it, as permitted by EUWA 2018, s.6(2).[39]

The threshold for the exercise of this jurisdiction is high. The only relevant factor is legal uncertainty, the interests of which must be, as Warby LJ notes, "so compelling", not merely "compelling". The strict standard of this test "reflects the key point, that any suspension represents a disapplication of legal rights which the legislature has conferred on natural or legal persons".[40]

The relevant principles for how to exercise this power can also be derived from the retained case law: (a) only temporary suspensions are possible; (b) the suspension should only last as long as is "strictly necessary" to ensure minimal interference with the normal legal order and the rights of those who would rely on the dominant legislation (*Gewestelijke*); (c) this does not mean a period of time that the Government would find politically or administratively convenient (the only factor that is considered is legal uncertainty, not political or administrative convenience); and (d) "[t]he court must be satisfied that the period of suspension is really needed, to avoid legal uncertainty."[41]

Warby LJ considered the initial approach of the Government to be "far too relaxed".[42] The standard for determining the duration of the suspension was that suspension was "really needed in the interests of legal certainty" for the "whole

[38] ibid [32].
[39] For these propositions, see ibid [32]–[33].
[40] ibid [32].
[41] ibid [33].
[42] ibid [41].

period of delay".[43] The Respondent was required to itemise the steps to be completed, giving estimates of time for them, to plead its proposed duration of the suspension. Despite misgivings over the Government's lack of urgency up to that point, Warby LJ accepted the Government's proposal to suspend the disapplication until 31 January 2022.[44] As a result of this collaborative approach, relatively little analysis in the judgment is dedicated to this issue.

The more analytically interesting issue is the Court's decision to exercise the jurisdiction in the first place. The threshold for the exercise of the jurisdiction is high—a position in the retained EU authorities, endorsed by Warby LJ. It is clear that the only relevant consideration is legal certainty. Without stating this, however, Warby LJ appears to consider two other factors. The first is the type of inconsistency. As in *Chester* and *Liberty*, reconstruction of a legislative scheme was necessary, "not complete destruction".[45] The reason for the finding of inconsistency was the absence of safeguards required by Article 23(2), rather than the derogations themselves.[46] Warby LJ appears to consider this context, along with the Government's stated intention to devise and implement the necessary safeguards, as relevant to the question of whether to exercise the jurisdiction. Going beyond strict analysis of legal certainty, Warby LJ considers the "serious practical difficulties" that immediate disapplication of the Immigration Exemption would cause for the Home Office.[47] Warby LJ relies on the findings in the Main Judgment, which show that the Immigration Exemption "has been and still is extensively relied on by the Home Office".[48] In light of this, Warby LJ finds that "the extent and significance of such disruption lends convincing support to the case for overriding, for a period of time, the substantive rights at issue".[49]

Respectfully, this conclusion is not consistent with the case law that Warby LJ endorses earlier in his analysis. These considerations are issues of political and administrative convenience. The case law is clear that the right of the individual to rely on their dominant EU rights may only be exceptionally overridden, and only in the interests of legal certainty. The thrust of the judgment's analysis is more concerned for the right of the Home Office to rely on the inconsistent provision of domestic law than the rights of individuals to rely on their dominant substantive rights. The purpose of the strictness of the test is to prevent suspensory relief from being used in this way. The focus on the interests of legal certainty, and the need for these interests to be "so" compelling, ensures that this is a narrow derogation

[43] ibid [41].
[44] ibid [46], [55].
[45] ibid [35].
[46] ibid [36].
[47] ibid [38].
[48] *Open Rights Group (No 1)* (n 7) [16]–[17].
[49] *Open Rights Group (No 2)* (n 6) [38].

from the ordinary rule. The scope of what constitutes "interests of legal certainty" requires careful definition to prevent the subversion of the legal policy of the rule. If it can extend to administrative inconvenience, the rule loses its narrow scope. Lord Gardiner LC's Practice Statement is useful on this point. In this statement, legal certainty embraced private law relationships such as "contracts, settlements of property and fiscal arrangements", as well as criminal liability. The concept of legal certainty is for the benefit of natural and legal persons, who should be able to conduct themselves without worrying that the underlying legal basis for their dealings and conduct might change without notice or ease of comprehension. It might be the case that immediate disapplication of the Immigration Exemption causes the Home Office great practical difficulties. But this is not the same as legal uncertainty. If anything, the legal position of the Home Office would have been quite clear, if unwelcome: until the Government introduced an Article 23 compliant Immigration Exemption, the Home Office would not be able to rely on it.

A stronger argument in favour of suspending relief is offered later in the judgment when Warby LJ rejects Ben Jaffey QC's suggested partial suspension of relief. Warby LJ observes that Parliament has "progressively imposed significant legal responsibilities on private sector actors, such as employers, landlords, and transport operators."[50] This policy is backed up, in places, by criminal sanctions for those who fail to discharge their duties.[51] Given the potential impact of immediate disapplication on private actors, such as employers and landlords, legal uncertainty is potentially a relevant concern. But none of this is certain. The strictness of the test necessarily demands careful analysis of affected legal relationships and the potential impact of disapplication. The unanswered question in the judgment is whether disapplication of the Immigration Exemption would cause landlords and employers any difficulties—let alone serious difficulties, as required by the high threshold of the test. Warby LJ holds that the Court can "reliably infer that a major shift in the law would cause significant disruption for the private sector".[52] Respectfully, the strictness of the test demands more than an inference. It cannot be possible, except in the rare truly self-explanatory situation, for the strict "serious difficulties" (*B v Latvia*) threshold to be cleared without analysis of the impact of the change on affected classes of persons. Without this, "legal uncertainty" becomes too impressionistic a concept to provide a disciplined restraint on the exercise of this jurisdiction.

Given all this, it may be that the Court of Appeal did in fact decide to exercise its power in Regulation 5 to depart from retained EU case law. This would

[50] ibid [49].
[51] For example, the Immigration Act 2014, s 33A.
[52] *Open Rights Group (No 2)* (n 6) [50].

be despite the Court's statement to the contrary. On close reading of Warby LJ's judgment, it appears that factors other than legal certainty may be considered by courts deciding whether to exercise the retained suspensory jurisdiction. One such factor is the type of inconsistency identified by the Court. Where the inconsistency is the absence of a safeguard, as in this case as well as in *Chester* and *Liberty*, the Court may be justified in suspending relief. Another factor appears to be administrative inconvenience, although only to those relying on the inconsistent law at the time of judgment and not to those tasked with devising and implementing a new law.[53] Warby LJ appears to have been influenced by the potential of immediate disapplication to disrupt the operations of the Home Office. It is possible that government departments may raise similar points in future hearings on the issue of suspending relief. Such considerations are, according to the strict test in the retained EU case law, not relevant. This case may be seen as setting a precedent for their relevance in future English and Welsh cases.

VII. Conclusion

The claimants in this case were two NGOs. One can only speculate how the absence of a specific "wronged" individual affected the course of proceedings. Perhaps the Court of Appeal might have been more reluctant to exercise the suspensory jurisdiction if it had been faced with a litigant who stood to lose personally from the decision not to make an immediate declaration. As it was, the only party personally affected by the decision was the Home Office. This might explain the unexpected attention given to what ought to have been, according to the retained case law, an irrelevant consideration.

The analysis underpinning this judgment points to a lower standard than the strict standard inherited from the CJEU (and endorsed on the face of the judgment). If the underlying logic of the legal policy is examined, this makes little sense. The rule arose in the multi-level jurisdiction of the EU. The potential for inconsistency between national and Union law did, and does, lead to concerns for legal certainty.[54] This suspensory jurisdiction was a tool to be used exceptionally to mitigate the most compelling instances of legal uncertainty caused by the principle of primacy, where immediate disapplication would cause "serious difficulties". On this analysis, the rule has a more modest role to play in the domestic UK jurisdiction outside the EU. As the UK legislates in areas that are currently served by retained EU law, the importance of this suspensory jurisdiction

[53] ibid [33].

[54] For example, issues arising from the "incidental effect" of directives. See Case C-194/94, *CIA Security International SA v Signalson SA and Securitel Sprl.* [1996] ECR I-2201, ECLI:EU:C:1996:172.

will decrease. The potential for uncertainty arising from the principle of primacy of EU law is much reduced, and likely to decline further. If the conditions for its exercise in the EU system were tight, they ought to be as tight or tighter in the UK jurisdiction. It would be contrary to principle, as affirmed in the judgment, for administrative inconvenience to become a relevant factor in deciding whether to suspend an individual's enjoyment of a dominant right.