Issue [5] 2016

OXFORD UNIVERSITY
UNDERGRADUATE LAW JOURNAL

OXFORD UNIVERSITY
UNDERGRADUATE LAW JOURNAL

-The-

5th Editorial Board of the Oxford University Undergraduate Law Journal

2015-2016

-The-
Honourary Board

Sir Nicolas Bratza

Professor Michael Bridge

Donald Findlay QC

Professor Christopher Forsyth

Ian Gatt QC

The Right Hon the Lord Judge

The Right Hon the Lord Kerr of Tonaghmore

Michael Mansfield QC

The Right Hon the Lord Neuberger of Abbotsbury

The Right Hon the Lord Phillips of Worth Matravers

Lord Pannick QC

His Honour Judge Gordon Risius CB

Dinah Rose QC

Sir Konrad Schiemann

The Right Hon the Lord Wilson of Culworth

CONTENTS

Foreword

Dr James Goudkamp, Fellow of Keble College, Associate Professor of Law

Most law journals and reviews in the common-law world are administered by student editorial committees. This is especially true in the case of Australia, Canada, and the United States.[1] The United Kingdom is an outlier in this regard, with the preponderance of journals and reviews being run by faculty members. The *Oxford University Undergraduate Law Journal* bucks this trend. It is student-run, with limited input and oversight coming from the Oxford Law Faculty. The *Journal* is distinctive in another respect: not only is it run by a student editorial team but its contributions are authored exclusively by undergraduate students. The Editors-in-Chief observe in their Introduction that this is the first issue of the *Journal* to include contributions penned by students from outside of Oxford.

Interesting debates can be had about the functions served by law journals and reviews.[2] There is similarly scope for stimulating argument about whether it is better for law journals and reviews to be edited by students or faculty members.[3] Finally, one can discuss fruitfully whether students should (and should be afforded a platform to) publish their work. It is themes of this nature, all of which are largely unexplored in the case of England, that I want briefly to explore in this Foreword. In doing so, I mean no disrespect to the contributors to the present issue (all of whom I am certain are among the most capable students in their cohort) or to the current editorial team. I simply wish to explore critically certain questions in the traditional academic spirit. My aim is to be constructive in doing so, and I hope that my remarks will be understood with that end in mind. Readers who are seeking an overview of the contents of the present issue are referred to Introduction written by the Editors-in-Chief.

I wish to begin by asking whether students should publish their work. There are several reasons why the opportunity to publish is (or should be) attractive to students (I encourage my own students to pursue it vigorously, especially my doctoral candidates). Most obviously, a student who publishes his or her work thereby marks him or herself out from nearly all of his or her peers. This can be an obvious advantage when it comes to applying for positions in law firms, pupillages, scholarships, and further study. Writing for journals and reviews can also be educative. Students are afforded an opportunity to develop thoughts that might otherwise remain embryonic. There is, therefore, a potential payoff in so far as performance in examinations is concerned.

However, there are also reasons why students might not wish to publish their work. Indeed, it might be suggested that students should be protected from themselves and that their work should not put into the public domain. The point might be developed in the

[1] The practice in the United States is thoroughly institutionalized and has by far the longest legacy. For exhaustive discussion, see ML Closen and RJ Dzielak, 'The History and Influence of the Law Review Institution' (1997) 30 Akron Law Review, Issue 1, Article 2.

[2] Consider MD Kirby, 'Welcome to Law Reviews' (2002) 26 Melbourne University Law Review 1.

[3] One provocative engagement with this theme is RA Posner, 'Law Reviews' (2006) 46 Washburn Law Journal 155.

following way. Students, because they are at the beginning of their legal careers, will naturally have a relatively incomplete view of the legal universe. Consequently, their contributions are bound to overlook important points or, at a minimum, be insufficiently nuanced in certain respects. One American scholar wrote trenchantly in 1926: '[s]tudent notes cannot build up a body of doctrine. They will always lack that breadth of legal knowledge and maturity of view which can only come to one who has lived with a specialty for many years.'[4] In these circumstances, is it really in students' interests to have their thoughts immortalized in print?

It might be said in response that this question (or at least a question similar to it) can properly be asked also in relation to established researchers. Thus, it is not uncommon for experienced writers to regret publishing at least some of their writing. The late Peter Birks famously remarked towards the end of his glittering career that '[a]lmost everything of mine now needs calling back for burning'.[5] Birks evidently wished that he had not put most of his thinking into the public domain. It might be, therefore, that law students are no more likely to regret (or have cause to regret) the publication of their work than legal academics.

Is it worth reading student-authored contributions? It might be thought that there are two reasons not to do so. The first reason is that student contributions might offer analyses that are underdeveloped or simply wrong. However, this reason has limited force. For one thing, there is often much to learn from the mistakes and failings of others. The study of error is frequently revealing. Furthermore, this first reason is not specific to student-written articles. The cold hard truth – indeed, it is a notorious fact – is that much published work is simply not worth reading. Academics are all but compelled to publish endless streams of text on account of large a variety of sticks and carrots.[6] It is hardly surprising, in these circumstances, that the quality of academic writings generally is highly variable. Accordingly, it is unlikely that there is anything unique in so far as quality is concerned about student-written work.

The second reason to avoid student-authored work is that students, because their ideas have had less time in which to mature, might be more prone than established writers to resile from their published positions. It is worth recalling in this regard that judges, for much of the history of the common law, declined to cite the work of living scholars, lest any scholar adjust their thinking.[7] This judicial convention, however, was difficult (if not impossible) to justify, and it is unsurprising that it has been abandoned.[8] It is very hard to see why the fact that a writer *might* change his or her stance is a reason for not to engage with the writer's work. Even if a given analysis is later rejected by its original architect, it does not follow that the analysis is defective. It may well be that it is the *abandonment* of the analysis that is unjustified rather than the analysis. This second reason (like the first) is unconvincing.

---

[4] A Kocourek, 'Editorial Note' (1926) 21 Illinois Law Review 147, 149.
[5] P Birks, *Unjust Enrichment* (2nd edn, OUP 2005) xii.
[6] For a highly entertaining yet searching and perceptive discussion, see L Green, 'Why every academic under 60 must have a blog' <https://ljmgreen.com/2016/09/14/why-every-academic-under-60-must-have-a-blog/> accessed 17 November 2016.
[7] See the remark of Kekewich J in *Union Bank v Munster* (1887) 37 Ch D 51, 54.
[8] For discussion, see A Braun, 'Burying the Living? The Citation of Legal Writings in English Courts' (2010) 58 American Journal of Comparative Law 27.

For the foregoing reasons, I very much doubt that there is any *a priori* reason to shun student-authored work. Without wishing to develop a fully formed position on the point, my view is that work should be read if the reader finds it profitable to do so. There is no basis, according to this criterion, to eschew student-written articles. I note that a number of student-authored works have had a significant impact in the United States.[9] There is no reason why the same could not and should not happen in the United Kingdom.

One final point is that it is notable that the current editorial committee of the *Journal* has refused to adhere rigidly to *The Oxford University Standard for Citation of Legal Authorities* ('***OSCOLA***'),[10] despite this being their professed style guide. For my part, I see no difficulty at all in this. Indeed, there is much to be said for refusing to impose on authors at least certain citation systems. Academics across the United Kingdom and beyond needlessly lose a great deal of time trying to divine and comply with the citation system that any given journal or review uses. Invariably, the 'guidance for contributors' that is given on the websites of journals and reviews raises many more questions than it answers. Most of the (very capable) postgraduate students whom I have engaged over the years to assist with editorial work have been flummoxed by style guides. *OSCOLA*, it is true, is nowhere near as oppressive and unwieldy as certain other citation guides. Indeed, it is mercifully economical compared with, for example, the grotesque *Australian Guide to Legal Citation*[11] and the *Bluebook*.[12] (The most recent edition of the latter runs to an astounding 560 pages.[13]) However, complying with it is nevertheless time-consuming.

In these circumstances, it is arguable that journals and reviews should permit authors to cite sources in any way that they wish subject to two conditions. The first is that sufficient information be imparted to permit the sources to which reference is made to be located by readers. The second is that the contribution concerned be (by and large) internally consistent in the way in which sources are cited; internal inconsistency risks misleading the reader to think that variations in the style are deliberate and thus meaningful. As well as relieving authors of a significant burden, relaxing the need for adherence to style guides is likely to reduce the amount of work that needs to be done by editorial committees, which will in turn cut down the amount of time between submitting an article and the publication thereof. The substantial time lag in this regard is one of the principal reasons why law journals and reviews are faced with an existential threat from alternative fora for publishing legal research. Effective solutions need to be found if journals and reviews are to remain relevant.

James Goudkamp
17 November 2016

---

[9] See E Volokh, *Academic Legal Writing: Law Review Articles, Student Notes, Seminar Papers, and Getting on Law Review* (3rd ed, Federation Press 2007) 5–6.
[10] *The Oxford University Standard for Citation of Legal Authorities* (4th edn, Hart Publishing 2012).
[11] *Australian Guide to Legal Citation* (3rd edn, Melbourne University Law Review 2010).
[12] *The Bluebook: A Uniform System of Citation* (20th edn, Harvard Law Review Association 2015).
[13] Richard Posner has mercilessly lampooned it, writing: 'Anthropologists use the word "hypertrophy" to describe the tendency of human beings to mindless elaboration of social practices. The pyramids in Egypt are the hypertrophy of burial. The hypertrophy of law is A Uniform System of Citation': RA Posner, 'Goodbye to the Bluebook' (1986) 53 University of Chicago Law Review 1343, 1343.

Introduction

Vincent Ooi and Denise Lim, Editors-in-Chief

It is with great pleasure that we introduce the 5th Edition of the Oxford University Undergraduate Law Journal. This Edition features a carefully curated selection of articles spanning a wide range of topics. As this is the first time that we have been able to accept submissions from outside Oxford University, we have been delighted to receive numerous well-written and insightful articles from around the United Kingdom. The most promising of these articles have been compiled into this Edition, which we hope will prove to be a most interesting read.

It is our firm belief that a desire and an ability to engage thoughtfully and productively in a subject should be the core of our education. Practical limitations in the undergraduate law course often mean scant opportunities for students to fully delve into the select areas of their study which might leave lingering queries in their minds. In light of this, the Journal seeks to complement our tutors' endeavours by providing a platform for students to engage with and propose new insights gleaned from their forays into the subject.

More often than not, students are inclined to viewing their subjects as clear-cut bodies of knowledge wholly and perfectly embodied in their neatly-bound textbooks and tomes. Yet, the common law itself is a living beast, a constantly evolving corpus of rules as renowned for its sagacity as it is for its peculiarities. It is for the students of our generation and future more, to join the discussion by being forthcoming in recognizing this and engaging in constructive dialogue with these rules. It is for this reason we so earnestly try to promote the journal to our peers and sponsors alike. We seek to emphasize that textbooks alone can offer no calm repose to settle legal arguments, and that opinions do not merely have to be neatly trimmed summaries of pre-established arguments only readied for timed assessment.

In light of this, the Journal remains at the forefront of current legal developments, with three case notes covering recent developments in Trust Law and Competition Law. Alexandra Clarke offers an engaging query on the nature of backwards tracing following the recent case of *Brazil v Durant*, while Sym Hunt examines the rightful role of rebates in Competition law in *Danmark*. Lastly, Matthew Hoyle provides an incisive analysis of *AIB v Mark Redler* and its impact on the law of breach of trust by wrongful disbursement.

Two articles draw attention to pertinent social issues. Vincent Ooi and Jia Wei Loh explore ways of ensuring that parents with disabilities are treated equally in Family Law. Azfer Ali Khan examines the Syrian crisis and considers the rights of refugees at International Law. These pieces serve as timely reminders that the law takes on the precarious role of being yet another response to particular visions of desired social and political theory, whilst simultaneously being the very scaffold which offers sustainability to theories thereof.

There were many strong contenders for the Norton Rose Fulbright Prize for the Best Article in Contract, Tort, Trusts, and Land Law this year. The articles were assessed on a blind review basis to guard against bias. Apart from the two abovementioned case notes on Trust Law, Ryan Chan-Wei's alternative framework of analysis for Rylands v Fletcher liability provides an interesting new take on the old rule. The Prize eventually went to Zi Xiang Tan for his daring proposal to introduce an independent doctrine of family homes proprietary estoppel. We are very grateful to Associate Professor Sandy Steel for assessing the articles and selecting the prize winner. In addition, we would like to thank Norton Rose Fulbright for their generous sponsorship of the Prize (since our 3rd Edition).

We sincerely hope that the Journal will continue to thrive in the coming years. Central to this is our enduring desire to maintain high standards of publication in our future issues. In light of this, each year's edition of the Journal is both a beginning and an end: a product of grit and painstaking measure, and an entreaty to potential writers and readers alike to consider supporting our humble enterprise.

The University has been supportive of our endeavours and the Faculty a valuable source of advice. We are particularly grateful to Associate Professors Sandy Steel, Rebecca Williams and James Goudkamp for their continued support of the Journal. Special thanks must go to our Editors for this Edition, Marius and Zi Xiang, for their painstaking work in ensuring the smooth publication of the Journal. Lastly, but surely not the least, our writers, associate editors and other board members have worked tirelessly on these articles to ready them for publication. It is only with their tireless work and eye for detail that we can offer this to you, and for that we will always be grateful.

Equity's New Child: The Birth of the Family Proprietary Estoppel

Zi Xiang Tan (Warren)*

Introduction

Much ink has been spilled in recent years on the common intention constructive trust ('**CICT**') and the proprietary estoppel. Judges have long noticed the similarities between the two doctrines[1] and commentators have long argued that the CICT can be considered a species belonging in the genus of proprietary estoppels.[2] In light of recent developments, this essay continues that line of work but with a new proposal: the family proprietary estoppel thesis.

Part 1 will first examine the doctrine of CICT, summarize the Gardner-Davidson theory and argue it is the best account of CICT. The Gardner-Davidson theory argues that the beneficial interest in the family home should be split according to the normative implications of the domestic relationship between its occupants. Part 1 will then argue that the doctrine of CICT, in its present state, is only able to accomplish the rationale posited by the Gardner-Davidson theory – namely giving effect to the normative implications of the parties' domestic relationship – in a limited manner.

Part 2 will examine the doctrine of proprietary estoppel and consider the domestic/commercial dichotomy devised by the most recent House of Lords case. It will then argue that the best way to understanding proprietary estoppel is to see it as a heterogeneous doctrine and within this heterogeneity, a subset of proprietary estoppels arising in a domestic setting will be identified. It will then discuss Birks' warning against unjustified remedial flexibility, which applies strongly to proprietary estoppel, before showing how his challenge can be met in principle and in the subset of proprietary estoppels identified earlier.

Part 3 will embark on the project of legal reform. It will first point out the burdens of proof taken up by any thesis seeking to assimilate CICT to proprietary estoppel, fully or partially. It will then evaluate three proposals: first, the conventional argument that the CICT should be judicially assimilated into proprietary estoppel ('**full assimilation thesis**'); second, the family proprietary estoppel thesis; and, third, that any reform should be undertaken by the Queen-in-Parliament ('**legislative reform**').

---

* University College, Oxford. I would like to express my heartfelt gratitude to Professor Simon Gardner and the associate editors at OUULJ for their invaluable feedback. The usual disclaimers apply.

[1] See e.g. per Browne-Wilkinson V-C in *Grant v Edwards* [1986] Ch 638, 656; per Lord Oliver in *Austin v Keele* (1987) 72 ALR 579, 609; per Nourse LJ in *Stokes v Anderson* [1991] 1 FLR 391; per Robert Walker LJ in *Yaxley v Gotts* [2000] Ch 162, 177; and per Chadwick LJ in *Oxley v Hiscock* [2005] EWCA Civ 546, [2005] Fam 211 [66].

[2] See e.g. D Hayton, 'Equitable rights of cohabitees' [1990] Conv 370, and subsequently in 'Constructive trusts of homes - a bold approach' (1993) 109 LQR 485; B McFarlane, *The Structure of Property Law* (Hart Publishing 2008), 775-782; and, more recently, YK Liew, 'The secondary-rights approach to the "common intention constructive trust"' [2015] Conv 210, 223.

1. *Doctrine of CICT and the Gardner-Davidson Theory*

Any statement of the law of CICT must begin with the leading decision by the Supreme Court in *Jones v Kernott* ('***Kernott***').[3] While *Kernott* has conclusively settled particular points of the law on CICT, it has failed to clarify how its conception of CICT as a whole differs from that of the past decisions of the House of Lords and Privy Council.[4] This has led Miles to claim that the 'contours [of the CICT] have arguably become no easier to discern' despite these cases.[5] Hence, the following is a provisional statement of this uncertain area of law.[6] Briefly, if:

(1) The defendant (D) owns a house; and
(2) The claimant (C) is a member of D's family,

Then:

(3) The proportions of C's and D's shares are *prima facie* determined by how the legal title is held:
    (a) If the legal title is held solely by D ('**sole name**'), then D has all the beneficial interest.
    (b) If the legal title is held jointly by C and D ('**joint names**'), then they each have 50% of the beneficial interest.

But:

(4) This *prima facie* position can be displaced by a *genuine* (i.e. actual and inferred) common intention by C and D that C should have a greater share in the ownership of the house; and
(5) The proportions of C's and D's shares are determined by:
    (a) their *genuine* common intention, if there is one; or
    (b) in the absence of a *genuine* common intention, the courts will impute one so as to achieve fairness between them.

(4) and (5) are also commonly referred to as the acquisition and the quantification of beneficial interests respectively. There are many outstanding controversies in this provisional account of CICT, including:

i. The scope of the CICT as determined by (1) and (2);[7]
ii. The difficulty in displacing the *prima facie* position in (3);
iii. The desirability of separating (4) and (5) as distinct analytical stages and how common intention

[3] [2011] UKSC 53.
[4] They are, in chronological order, *Pettitt v Pettitt* [1970] AC 777 (HL), *Gissing v Gissing* [1970] UKHL 3, *Lloyds Bank v Rosset* [1990] UKHL 14, and *Stack v Dowden* [2007] UKHL 17 and *Abbott v Abbott* [2007] UKPC 53.
[5] J Miles, 'Cohabitation: Lessons for the South from North of the Border?' [2012] CLJ 492.
[6] The analysis in *Kernott* applies to both joint names and sole name cases, following [51] to [52]. Even this is a controversial point. The failure of *Stack* and *Kernott* to explicitly overrule *Rosset*, a sole name case that laid down detrimental reliance by C as a requirement, means that one could plausibly argue that the law currently treats joint names and sole name cases differently. However, such a distinction cannot be justified. The better view is that one should not read too much into whether the property is held jointly or solely in the context of family homes, this being more often than not a historical contingency than an act intended to have legal significance.
[7] What types of relationship between C and D and what kinds of purpose for which the home was bought and used will trigger the trust? The Court of Appeal in *Laskar v Laskar* [2008] EWCA Civ 347 seems to rule out a CICT in a case of 'investment for rental income and capital appreciation, even where [the parties'] relationship is a familial one'. But reliance on *Laskar* ought to be tempered by the realization that Lord Neuberger is applying his dissenting analysis in *Stack*; cf *Geary v Rankine* [2012] EWCA Civ 555, *Aspden v Elvy* [2012] EWHC 1387 (Ch), and *Ullah v Ullah* [2013] EWHC 2296 (Ch) for more variation on this theme.

is inferred, implied, or imputed, and how, if at all, these methods of ascertaining intention differ;

iv. The meaning of fairness in (5); and

v. Whether the requirement of detrimental reliance by C on her common intention with D, per *Lloyds Bank v Rosset*,[8] has survived *Stack v Dowden* ('***Stack***')[9] and *Kernott*.[10]

The family proprietary estoppel thesis is built upon the Gardner-Davidson theory, which is not neutral on these controversies. It is founded on the view that the development of the CICT is part of the increasing 'familialization' of property law[11] and that the doctrinal elements of CICT must be thus understood. In a domestic context where individuals repose trust in each other, it is neither sensible to expect nor reasonable to require adherence to formality rules. This justifies the informal creation of equitable interests under a CICT. Similarly, certainty is less important as people in family relationships are less likely to dispose property with an eye on the relevant law[12] and discretion is necessary to effect justice in family matters. All these stand in stark contrast to the norms in a commercial paradigm, where adversarial parties deal with each other at arm's length and legal certainty is paramount.

According to this theory, the appellate courts do not take the requirement of genuine common intention seriously.[13] This is best illustrated by *Kernott*. In that case, while the Supreme Court unanimously discerned a genuine common intention in (4) – a point which was itself based on the trial judge's finding that the Court of Appeal had described as without evidential foundation[14]—they split 3-2 on (5). The majority went on to find a genuine common intention as to the shares,[15] whereas the minority disagreed and imputed a common intention on the basis of fairness.[16] Magically, they arrived at the same distribution of 90-10. This divergence in approach and eventual convergence in outcome are unlikely to be compatible with taking seriously the requirement of genuine common intention.

---

[8] [1990] UKHL 14.

[9] [2007] UKHL 17.

[10] Conceivably, an essay arguing for proprietary estoppel and CICT to be (fully) assimilated might see some significance in detrimental reliance appearing as an element of proprietary estoppel and possibly CICT. This is a mistake. Rather, this essay argues that the normative implications of the relationship are the primary driver. For a contrary view, see B Sloan, 'Keeping Up With the *Jones* case: Establishing Constructive Trusts in "Sole Legal Owner" Scenarios' (2015) 35 LS 226, which argues that the requirement of detrimental reliance persists in CICT, and is crucial to justifying equity's intervention.

[11] It is the judicial process whereby individualistic, relationship-neutral property law principles affecting ownership of the family home are reinterpreted and recalibrated to accommodate the specific needs of family members. This was first identified in J Dewar, 'Land, Law, and the Family Home', in S Bright and J Dewar (eds), *Land Law. Themes and Perspectives* (OUP 1998).

[12] As Waite LJ observed in *Midland Bank plc v Cooke* [1995] 4 All ER 562, 575: 'When people, especially young people, agree to share their lives in joint homes they do so on a basis of mutual trust and in the expectation that their relationship will endure. Despite the efforts that have been made by many responsible bodies to counsel prospective cohabitants as to the risks of taking shared interests in property without legal advice, it is unrealistic to expect that advice to be followed on a universal scale. For a couple embarking on a serious relationship, discussion of the terms to apply at parting is almost a contradiction of the shared hopes that have brought them together. There will inevitably be numerous couples, married or unmarried, who have no discussion about ownership and who, perhaps advisedly, make no agreement about it. It would be anomalous, against that background, to create a range of home-buyers who were beyond the pale of equity's assistance in formulating a fair presumed basis for the sharing of beneficial title, simply because they had been honest enough to admit that they never gave ownership a thought or reached any agreement about it.'

[13] S Gardner, 'Problems in Family Property' [2013] CLJ 301, 306-307.

[14] [2010] EWCA Civ 578, [81]-[83].

[15] *Kernott* (n 3) [48]-[49]. This is even though the trial judge himself had not done so.

[16] ibid [76]-[77].

As such, the CICT does not truly respond to common intention. From a legal realist perspective, purporting to respond to common intention serves to legitimize the outcome by dressing up the reasoning in the libertarian clothing that pervades other areas of private law. This is regrettable as it will cause difficulty for judges who apply the law and practitioners who advise their clients. Rather, it should be openly acknowledged that the CICT responds to (or ought to respond to) the normative implications of the parties' domestic relationship ('**NIDR**') instead.

The class of relationships capable in themselves of bearing normative implications cannot be exhaustively catalogued. Classic examples include cohabiting couples in CICT and inheritance representation cases in proprietary estoppels: where C has been led to believe that C would be granted property on death, if not before, and, in that belief, C has acted to their detriment by providing services to D, and sometimes as suffering other detriment.[17] To provide a non-exhaustive illustration, broad principles will be outlined with respect to cohabiting couples in CICT. As the ultimate goal of the CICT is asset redistribution, the Gardner-Davidson theory focuses on the *material* implications of the relationship. For example, do they pool their resources across the board? If the relationship is materially communal,[18] then the maxim 'equity is equality' is the starting point: for example, between two economically equal cohabitants, each should receive equal shares in the family home. However, there is no categorical imperative for strict equality: in *Kernott*, the eventual 90-10 split in favour of C arguably displays a new, more refined vision of 'equality': as D owned another house, this distribution meant she took roughly half of the combined value of the two houses. Furthermore, if there are dependants or disabilities on the part of either partner, this presumably should be taken into account. If the relationship is materially non-communal,[19] their shares are determined by reference to their contributions,[20] probably on the basis of reversing C's unjust enrichment of D.[21] But such an approach should be tempered by a sensitivity to the facts: for instance, if C and D have vastly disparate earning power and do not pool their resources at the insistence of the more powerful D, the appearance of material non-communality should not stop the judge from effecting a more humane and just outcome. These are matters of open-textured judgment that cannot be prescribed definitively in advance.

In truth, the Gardner-Davidson theory, on which the family proprietary thesis is based, is but

---

[17] A list of such cases in recent years: *Re Basham* [1986] 1 WLR 1498 (Ch); *Wayling v. Jones* (1993) 69 P&CR 170 (CA); *Walton v Walton* Unreported April 14, 1994; *Gillett v. Holt* [2001] Ch 210; *Jennings v. Rice* [2002] EWCA Civ 159; *Campbell v. Griffin* [2001] EWCA Civ 990; *Grundy v. Ottey* [2003] EWCA Civ 1176; and *Uglow v. Uglow* [2004] EWCA Civ 987. It is central to the family proprietary estoppel thesis that commercial examples like *Crabb v Arun DC* [1976] Ch 179, *AG for Hong Kong v Humphreys Estate Ltd* [1987] 2 All ER 387 (PC), and *Cobbe v Yeoman's Row* [2008] UKHL 55 are not normatively similar enough to be assimilated.

[18] Examples in case law include *Abbott v Abbott* [2007] UKPC 53, *Bank of Scotland v Brogan* [2012] NICh 21 [42], and *Kernott* (n 3).

[19] *Stack* is an example, in which C and D were unmarried and had kept their financial affairs very separate and C had contributed (at least) 65% of the cost of acquiring the house and received 65% of the ownership of the house.

[20] There is continuing controversy over which contributions are relevant. It is arbitrary to look at who happens to make mortgage payments. cf *Stack* to *Aspden v Elvy* [2012] EWHC 1387 (Ch) and *Thompson v Hurst* [2012] EWCA Civ 1752, which seem to go beyond *Stack* without properly confronting the House of Lords precedent. Practically speaking, there are difficult issues of quantification within the family context which the law cannot deal with precisely, necessitating the use of discretion.

[21] See S Gardner, 'Family Property Today', (2008) 124 LQR 422, 437-439, for the difficulties in doing so. Etherton LJ's ambitious thesis in 'Constructive Trusts: A New Model for Equity and Unjust Enrichment' [2008] CLJ 265 was that *Stack* is an example of a remedial constructive trust granted in response to a new ground of restitution for autonomous or subtractive unjust enrichment. This view has not gained currency and will not be considered further, except to note that the challenge of discretionary remedialism is more easily met in the family context, rather than within the emerging and unsettled law of restitution.

another stand in this controversy. While it has been cited by the Supreme Court,[22] it has never been fully adopted without reservation. As such, Gardner himself concedes the possibility that 'a judge could wrong-foot [him] on this tomorrow, by giving the law a trajectory different from that which [he] claim[s] to discern'.[23] This has arguably already happened in *Graham-York v York*.[24] However, the reason for this is 'the lack of an authoritative 360° exegesis (…) for in its absence a judge could not unreasonably strike out in a number of different directions'.[25] As such, even if the Gardner-Davidson theory is wrong, there must be an alternative exegesis of CICT. If a successful account of CICT is not possible, then it must be excised from English private law. With this disclaimer in mind, it is submitted that the family proprietary estoppel thesis is nonetheless viable because the Gardner-Davidson theory remains the best exegesis of the law.

At this juncture, having summarized the Gardner-Davidson theory and its place in the ongoing controversy, it is helpful to note one of its implications: that C's rights in a CICT are necessarily indeterminate until they are crystallized in a judgment, i.e. there is a certain level of judicial discretion in ordering relief. This will be explicitly justified in Part 3. The arguments that apply to proprietary estoppels that effectuate NIDR apply with equal force to CICT. As C's rights are indeterminate and subject to judicial discretion, it has understandably led some to compare these rights to remedial constructive trusts.[26] But the difficulty with this view lies in the fact that there is nothing approaching a consensus on what 'institutional' and 'remedial' mean in the first place, or indeed whether all constructive trusts have to be one or the other.[27] Given that the idea that English law currently does (or should) recognize remedial constructive trusts (whether in general or CICT in particular) has met with firm judicial[28] and extra-judicial[29] rejection, characterizing C's rights this way is unnecessarily controversial.

For our purposes, it is important to note that the libertarian characterisation of C's rights – as determinate, involving no judicial discretion and responding to the parties' common intention – requires accounting for when C's rights arose and how they changed. In a joint names scenario where there is no express declaration as to the nature or proportion of co-ownership and it is not shown that the parties had a different common intention at the time when they had acquired the property,[30] English law's presumption of an equitable joint tenancy is not rebutted. This presents two problems to the libertarian characterisation.

Firstly, if C's equitable interests are determinate and distinct from those of D's by the time the case

---

[22] *Kernott* (n 3) [21], [24].
[23] Gardner, 'Problems in Family Property' (n 13) 312.
[24] [2015] EWCA Civ 7.
[25] Gardner, 'Problems in Family Property' (n 13) 304.
[26] See, e.g. Etherton LJ, 'Constructive trusts and proprietary estoppel: the search for clarity and principle' [2009] Conv 104.
[27] This was helpfully pointed out by Professor Simon Gardner.
[28] See e.g. *Re Polly Peck International Plc (No 2)* [1998] 3 All E.R. 812 (CA), 823–825, per Millett LJ; and 830–832, per Nourse LJ, *Satnam Investments Ltd v. Dunlop Heywood & Co Ltd* [1999] 3 All ER 652 (CA), 670, per Nourse LJ, and *Crossco No 4 v Jolan Ltd* [2011] EWCA Civ 1619 [84], per Etherton LJ.
[29] Lord Neuberger, 'The Remedial Constructive Trust – Fact or Fiction' (Speech at the Banking Services and Finance Law Association Conference, Queenstown, 10 Aug 2014) <https://www.supremecourt.uk/docs/speech-140810.pdf> accessed 21 March 2016. Also of note are more radical voices at the margin that see all constructive trusts today as either misclassified express trusts or merely a remedy given by a court. See W Swadling, 'The Common Intention Trust in the House of Lords: an opportunity missed' (2007) 123 LQR 511, for a flavour of this view in the context of CICT and W Swadling, 'The Fiction of the Constructive Trust' (2011) 64 CLP 399 more generally.
[30] *Goodman v Gallant* [1986] Fam 106 (CA) and *Malayan Credit Ltd v Jack Chia -MPH Ltd* [1986] AC 549.

is litigated, there must be a non-discretionary event that severed their joint tenancy. As Dixon notes, this is 'one lingering and unremarked issue in both *Stack* and *Kernott*'.[31] Liew argues that the CICT does not reflect any of the established methods of severance[32] and the only remaining plausible explanation is that severance occurs by judicial discretion.[33] While this does not mean C's rights are necessarily indeterminate, it implies that, in at least this respect, judicial discretion has replaced hard rules. This fits the Gardner-Davidson theory better than the general property law.

Secondly, even if the parties' initial beneficial joint tenancy had subsequently been severed in accordance with an established method of severance, the parties would have a tenancy in common with equal shares. As Mee pointed out, the ambulatory nature of this tenancy in common introduces another complication to the libertarian characterisation.[34] Given that the proportions of the interest may change from 50-50 to a new distribution depending on their common intention, this would *prima facie* fall afoul of s 53(1)(c) of the Law of Property Act 1925, which requires that any disposition of a subsisting equitable interest or trust (i.e. equitable interest responding directly to intention) must be in writing, a point raised by Walton J.[35] Mee's solution is that each new division of the beneficial ownership occurs under a 'newly refreshed' CICT, which is caught by the exemption in respect of "the creation or operation" of non-express trusts in s 53(2) of the 1925 Act. Such a contrived view is not out of place within the convoluted and much-criticized jurisprudence of s 53(1)(c).[36] However, it is submitted that the better view is simply that, following the Gardner-Davidson theory, the CICT, owing to its domestic context, is a disorganized mode of creating rights *in rem* which are only determinate on judgment day.[37]

But if the Gardner-Davidson theory is correct, its central rationale – giving effect to NIDR – has implications on how the doctrinal elements of CICT should be interpreted. Firstly, (1) is questionable: in principle, the availability of remedy for C should not turn on D's ownership of the house. Whether C and D are long-time licensees, leaseholders, or freeholders has no bearing on NIDR. But under CICT, courts are limited by its inability to order personal remedies. Secondly, the theory would mean that 'family' in (2) should not be not construed too strictly. Indeed, in *Gallarotti v Sebastianelli*,[38] the CICT rules were applied to friends in a platonic relationship who decided to live together and went into business. It is easy to imagine countless permutations of domestic relationships involving mutual trust whose normative implications include asset redistribution. Thirdly, the *prima facie* position in (3) must be easily displaced as how the property is held is only weakly correlated to NIDR. Fourthly, as argued above, the courts' expansive approach to genuine common intention undercuts the libertarian nature of (4) and (5)(a), which is at odds with CICT's rationale and thus better seen as fig leaves that camouflage CICT within the rest of property law. The work of quantifying beneficial interests is almost wholly done by (5)(b). As such, the doctrinal elements of the CICT must be re-interpreted. Part 3 will show the result of this re-interpretation: the family proprietary estoppel.

---

[31] M Dixon, 'The Still Not Ended, Never-ending Story' [2012] Conv 83, 84.
[32] These methods consist of *inter alia* written notice, under Law of Property Act 1925 s.36(2), operating on one's share, mutual agreement, course of dealing, or the murder of a joint tenant, under Forfeiture Act 1982 s 1.
[33] Liew (n 2) 218.
[34] J Mee, 'Ambulation, severance, and the common intention constructive trust' (2012) 128 LQR 500.
[35] *Richards v Dove* [1974] 1 All E.R. 888, 894.
[36] See S Gardner, *An Introduction to the Law of Trusts*, (3rd edn, OUP 2011), 105-106.
[37] See n 74 and its accompanying text.
[38] [2012] EWCA Civ 865.

2. Proprietary Estoppel as a Heterogeneous Doctrine

Any doctrinal statement of proprietary estoppel cannot be 'both comprehensive and uncontroversial' as there is uncertainty about virtually all of its aspects.[39] The standard account states that C has a claim against D in proprietary estoppel if:

(1') D has committed an act of assurance or acquiescence;
(2') D's act leads C to reasonably believes that C has or is going to get a right in D's property (which must be land, on some accounts);
(3') C acts in detrimental reliance on this belief; and
(4') It would be unconscionable to leave matters as they stand.

The House of Lords decisions of *Cobbe v Yeoman's Row Management Ltd* ('***Yeoman's Row***')[40] and *Thorner v Major* ('***Thorner***')[41] illustrate the instability of the law on proprietary estoppel. The polarity of the two cases has led commentators to announce, in a manner akin to Jesus Christ,[42] that they spelt the death[43] and the resurrection[44] of proprietary estoppel respectively.

In the later decision that is *Thorner*, Lord Neuberger was aware of the need to reconcile its outcome with *Yeoman's Row*. To that end, he did this both doctrinally and using the domestic/commercial dichotomy.[45] For our purposes, the doctrinal approach need not detain us.[46] It is Lord Neuberger's dichotomy that is relevant for our purposes: he noted that *Yeoman's Row* was a *commercial* dispute between two property developers who are presumably well-acquainted with the cautionary and channelling functions of formalities in land law,[47] whereas *Thorner* was but the latest iteration in a long series of inheritance representation cases arising within the *domestic* context and reflecting the court's sympathy for C in the form of less exacting judicial standards.

This dictum – while helpful to the family proprietary estoppel thesis – has come under criticism by commentators. For a start, Mee thinks that the normative positions of Mr Cobbe and David Thorner are closer than we might think. Adopting Goymour's terminology,[48] Mee argues that whereas that Mr Cobbe (in *Yeoman's Row*, a commercial example of the imperfect promises case) was a 'commercial risk-taker',

---

[39] S Gardner, *An Introduction to Land Law*, (4th edn, Hart Publishing 2015), 143.
[40] [2008] UKHL 55.
[41] [2009] UKHL 18.
[42] Coincidentally, the first draft of this essay is written during Holy Week 2016.
[43] B McFarlane and A Robertson, 'The Death of Proprietary Estoppel: Yeoman's Row Management Ltd. v. Cobbe' [2008] LMCLQ 449.
[44] J Uguccioni, 'The Resurrection of Proprietary Estoppel' [2009] LMCLQ 436.
[45] [2009] UKHL 18, [93].
[46] It is the proposition that, whereas a representation which was clearly made to be revocable could not be the basis of a proprietary estoppel claim, the converse was not true. In *Yeoman's Row*, C knew the commercial agreement was binding in honour only and that both parties could walk out of negotiations without any legal liability; and that is why he could not reasonably rely on D's oral assurances. In *Thorner*, while representations about testamentary intent are inherently revocable, insofar as C's reliance is independently reasonable, he does not need to prove that that he believed that D was legally bound to leave him the farm. For an elaboration on this, see Uguccioni (n 44) 440. Mee shares a similar view in 'The Limits of Proprietary Estoppel: Thorner v Major' [2009] CFLQ 367, 374.
[47] See L Fuller, 'Consideration and Form' (1941) 41 Columbia Law Review 799 for a full exposition on formalities.
[48] A Goymour, 'Cobbling Together Claims where a Contract Fails to Materialise' [2009] CLJ 37. For a discussion of officiousness and risk-taking in the context of unjust enrichment, see e.g. G Virgo, *The Principles of the Law of Restitution* (2nd edn, OUP 2006), 39-40.

David was a 'domestic risk-taker'.[49] This has been rightly criticized as '[reflecting] unrealistic expectations about the options open to a hard-working claimant in a domestic scenario'.[50] Furthermore, Mee argues that the factual boundary 'is far from secure'[51] and Sloan cautions against proprietary estoppel going the way of the CICT.[52] This point will be returned to in Part 3, where the dichotomy will be considered more fully.

It is submitted that the doctrinal approach to reconciling *Yeoman's Cobbe* and *Thorner* is unhelpful as it obscures the true nature of proprietary estoppel by presenting a unified doctrinal basis to what is really a heterogeneous creature. Unlike the CICT, which is relatively modern and whose application is confined to the context of family homes, proprietary estoppel dates back to the 19th century and is found in a wide range of situations. Within this diversity, commentators have noted that, despite commonalities among the three at an abstract level, it is practically more useful to discern three broad categories within proprietary estoppel.[53] They are:

i. Unilateral mistake cases: proprietary estoppel is classically applied to prevent landowners (D) from unconscionably asserting their strict legal entitlement as against another (C) who mistakenly thinks he owns the land and improves its value.[54]
ii. Common expectation cases: this category is distinguished from the first on the basis that here D has made a representation of an existing state of affairs to C, whereas D merely acquiesces in the first category.[55]
iii. Imperfect promise cases: where D promises (in an expansive sense of the word) a right *in rem* to C but fails to perfect it. Under certain circumstances, C can claim in proprietary estoppel. This category has developed the most in recent years and, unlike the previous two categories, it appears that C himself does not need to be a landowner.

Both *Yeoman's Row* and *Thorner* fall into the third category. It is submitted that Lord Neuberger's domestic/commercial dichotomy can be applied to this category, within which inheritance representation cases like *Thorner* are a further subset. In the language of the Gardner-Davison theory, in such an informal domestic context, proprietary estoppel effectuates the NIDR that C ought to receive something. While inheritance representation cases might be the most common subset of proprietary estoppel arising in a domestic context to be subsumed by the proposed family proprietary estoppel, it is submitted that in principle it is not the only one.[56]

The preceding account of proprietary estoppel omitted its controversial remedial flexibility.[57] The

---

[49] Mee (n 46) 374.
[50] B Sloan (2010) 22 SAcLJ 110, 129.
[51] Mee (n 46) 374.
[52] Sloan (n 50).
[53] See B McFarlane, *The Law of Proprietary Estoppel* (OUP 2014) ch. 1, Low, 'Nonfeasance in Equity' (2012) 128 LQR 63, 67, and Mee 'Proprietary Estoppel, Promises and Mistaken Belief' in S Bright (ed), *Modern Studies in Property Law, Vol. 6*, (Hart Publishing 2011), 182.
[54] E.g. *Ramsden v Dyson* [1866] LR 1 HL 129 (HL), and *Plimmer v Mayor, Councillors and Citizens of the City of Wellington* (1884) 9 App Cas 699 (PC).
[55] E.g. Lord Evershed MR's analysis in *Hopgood v Brown* [1955] EWCA Civ 7, [223].
[56] See e.g. the recent *Ely v Robson* [2016] EWCA Civ 774, in which it was unclear whether CICT or proprietary estoppel was being applied. As a rule of thumb, any case in which proprietary estoppel and CICT could plausibly be pleaded concurrently is caught by the family proprietary estoppel.
[57] There are many judicial acknowledgements of the remedial flexibility in proprietary estoppel. See e.g. *Ramsden v Dyson* (1866) LR 1 HL 129 (HL); *Plimmer v Wellington Corp* (1884) 9 App Cas 699 (PC); *Crabb v Arun DC*

mode of relief is less contentious: while it can operate both *in personam* and *in rem*,[58] this discretion is largely explicable on the basis of whether it is apt for C and D to live together. But the quantum of relief in proprietary estoppel is highly contentious. No single measure (i.e. C's expectation, detriment or restitutionary interest) is consistently pursued.[59]

It is submitted that this is principally due to the imperfect promise cases. Unilateral mistake cases and common expectation cases are more akin to the larger genus of estoppels and their remedies have a natural shape: D is simply estopped from making his claim. It is tempting to explain the remedial irregularity in imperfect promise cases as a matter of judicial discretion made in pursuit of fairness. But this remedial flexibility requires justification. In an essay on the general phenomenon of 'discretionary remedialism', Birks points out that separating liability and remedy in such an opaque manner is problematic. With characteristic eloquence, he argued that it would: firstly, 'make the management of litigation impossible, promoting unjust settlements based on guesswork as to the operation of the discretion'; secondly, 'deprive citizens of their dignity, bringing them as child-like supplicants (…) before a court which had grown much too big for its boots'; and, thirdly, 'threaten the stability of our society, because it would destroy the legitimacy of judicial authority' in a pluralist democracy, by failing to attain of 'transparent rationality [that] can be accepted by all sub-communities'.[60]

Fortunately, Birks' challenge can be met in principle. Gardner argues that the discretionary relief in proprietary estoppel is compatible with a democratic legal system and a commitment to the Rule of Law if three conditions of 'transparent rationality' – to adopt Birks' phrase – are met:[61]

i. The aim of the discretion must be fixed by the law;
ii. The discretion must be necessary; and
iii. Decisions taken under the discretion must be susceptible to audit.

While most judges have left it quite unclear how the discretion in proprietary estoppel operates, Walker LJ – as he then was – in *Jennings v Rice*[62] shed some light on this by arguing that the discretionary relief, firstly, must always be 'proportionate'[63] and, secondly, should reflect various other factors where these appear in the circumstances of the case. But Gardner points out that proportionality is about balancing ends, and not the end in itself. Thus an end must first be identified. Gardner concluded on a mixed note: the law as it stood 'cannot be sufficiently reconciled with the Rule of Law' but was 'fully capable of repair', with 'Walker LJ's analysis representing a promising start on this project'.[64]

It is submitted that the three conditions of transparent rationality are met where proprietary estoppel effectuates NIDR. Of necessity, the following arguments are borrowed from the Gardner-Davidson theory, in the similar context of CICT. Firstly, the aim of such a discretion is *ex hypothesi* fixed by the law: to effectuate NIDR. Secondly, the discretion is necessary. Consider the example postulated by Lord Scott in *Thorner*: what if D, before his death, needed to sell the farm in order to fund the costs of

---

[1976] Ch 179; *Pascoe v Turner* [1979] 1 WLR 431 (CA); *Sledmore v Dalby* (1996) 72 P & CR 196 (CA); *Jennings v Rice* [2002] EWCA Civ 159.

[58] Unlike the CICT, which, as the name implies, is confined to a right *in rem*.

[59] Gardner, *An Introduction to Land Law* (n 39) 153-154.

[60] P Birks, 'Rights, wrongs, and remedies' (2000) 20 OJLS 1, 23.

[61] S Gardner, 'The Remedial Discretion in Proprietary Estoppel – Again' (2006) 122 LQR 492, 505.

[62] [2002] EWCA Civ 159.

[63] See *Henry v Henry* [2010] UKPC 3, [65] for a similar sentiment: 'proportionality lies at the heart of proprietary estoppel and permeates its every application'.

[64] Gardner, 'The Remedial Discretion in Proprietary Estoppel – Again' (n 61) 512.

his medical treatment?[65] As Lord Scott pointed out, C had no strict entitlement to the farm and their relationship would imply that it is only fair for C to receive a portion of the proceeds of sale. If D's death occurred much sooner than either D or C anticipated, it would appear contrary to the implications of their relationship to award the entire farm to C. All these examples speak to the broader point that discretionary relief is necessary to effecting justice in family matters. Thirdly, such judicial decisions are susceptible to audit. There are manifestly unjust outcomes that essentially rule themselves out. But this point should not be overstated: the necessity of discretion suggests that this auditing is by nature open-textured and not mechanical. Within an intermediate range of reasonable outcomes, when all is said and done, we must rely on the judges' common-sense to do justice.

It is unnecessary for the family proprietary estoppel thesis to put the whole of proprietary estoppel on a footing of transparent rationality. Indeed, if proprietary estoppel is a heterogeneous doctrine, its remedial flexibility can only be justified within its particular context, as we have just done. The search for an overarching justification is futile. This is a point overlooked by proponents of the full assimilation thesis. The near-identical ways in which remedial flexibility is justified in CICT and proprietary estoppel paves the way for the family proprietary estoppel to be carved out from proprietary estoppel in general. This will bring greater conceptual rigour and result in a clearer guide for the relevant factors judges should look to in adjudicating such cases. Indeed, it is plausible that judicial recognition of the family proprietary estoppel could place implicit pressure on the residual areas of proprietary estoppel to meet the demands of transparent rationality.

3. *Two Weddings and an Act of Parliament*

As mentioned, three proposals for legal reform will be considered: the full assimilation thesis, the family proprietary estoppel thesis and legislative reform. It is necessary to point out the burdens of proof taken up by the first two proposals.

Firstly, they presuppose the legitimacy of proprietary estoppel, fully or in part. This is open to question,[66] particularly in light of its remedial flexibility. Besides justifying its remedial flexibility, which Part 2 has done in respect of the family proprietary estoppel thesis, it also requires proposing solutions to the following theoretical debates:

i. Justifying the circumvention of formality requirements relating to wills and land;
ii. The consequences of conditionality and the realization of the property forming the subject-matter of the representation;
iii. The significance of unconscionability as an independent element of the doctrine.

If Part 2 is correct, i.e. proprietary estoppel should be viewed as a heterogeneous doctrine, then there are no general correct answers to these debates. But even if proprietary estoppel is a monolithic doctrine, the unlikelihood of conclusively settling these debates counts against the plausibility of the full assimilation thesis. These challenges are watered down for the family proprietary estoppel thesis, as its scope is limited to the domestic context.

Secondly, both theses presuppose that the judiciary is well-suited to reform the law in this area. This is not without controversy.[67] On a thin account, any development of the common law must be made

[65] [2009] UKHL 18, [19].
[66] Sloan (n 50) 129.
[67] See J Mee, 'Burns v Burns: The Villain of the Piece', in S Gilmore, J Herring and R Probert (eds), *Landmark Cases in Family Law* (Hart Publishing 2010), 175 and 187, regarding democratic legitimacy in the judicial reform of

with reference to the courts' institutional role and the retrospective nature of common law judgments. On a thick account, the judiciary has a duty to develop the law in pursuit of a more coherent conception of justice.[68]

The conclusions can be summarized at this point. Firstly, the full assimilation thesis, while promising, fails because of the overwhelming difficulty of rationalizing the entire law of proprietary estoppel. Secondly, the family proprietary estoppel thesis redeems the practical benefits of assimilating CICT in a principled manner and proves to be the best proposal for legal reform. Thirdly, there is simply no feasible proposal for legislative reform in the horizon and any proposal must be compatible with the UK's commitments under the European Convention of Human Rights.

A. Full Assimilation: Left at the Altar

This essay began by noting that commentators have long argued in favour of the full assimilation thesis. But the enormity of the task of principled assimilation has never been fully confronted. It is apt to note that whatever initial judicial enthusiasm for this idea[69] had since met with a cold reception in *Stack*, the latest judicial word on this issue. Lord Walker opined in *obiter* that the CICT cannot be seen as a specific application of proprietary estoppel principles.[70] However, the two reasons given by Lord Walker are unsatisfactory.

His first reason is that assimilation will lead to the CICT losing its retrospectivity, and by extension its ability to bind a third party (X). With respect, this is wrong. Prior to the Land Registration Act 2002, there was legitimate doubt over whether a claim in proprietary estoppel was capable of binding third parties.[71] It is submitted that the orthodox view is that it could not: this policy consideration might explain why the courts have 'reinvented the proprietary estoppel in the guise of the [CICT]',[72] making CICT 'historically the court's preferred legal mechanism for giving effect to the property claims of a spouse or partner who had neither legal title nor entitlement under an express trust'.[73] However, the enactment of s 116(a) of the Land Registration Act 2002 ensures that by the time *Stack* was decided, C's equity by estoppel is capable of having proprietary effect against X, rendering Lord Walker's statement legally inaccurate.[74]

---

trusts and also Law Commission, 'Cohabitation: the Financial Consequences of Relationship Breakdown' (Law Com No 307 Cm 7182, 2007), paras 5.98-5.101.

[68] Dworkin, *Law's Empire* (Harvard University Press 1988), ch 8.

[69] Chadwick LJ in *Oxley v Hiscock* [2005] EWCA Civ 546, [2005] Fam 211 [66]: 'it seems to me very difficult to avoid the conclusion that an analysis in terms of proprietary estoppel will, necessarily, lead to the same result [as under a CICT analysis].' See also *Grant v Edwards* [1986] Ch. 638, 656; *Re Basham* [1986] 1 W.L.R. 1498 (Ch) at 1504; *Austin v Keele* (1987) AJLR 605 (PC), 609; *Lloyds Bank v Rosset* [1990] UKHL 14, [1991] 1 A.C. 107, 132-133; *Yaxley v Gotts* [2000] Ch 162, 176-177.

[70] [2007] UKHL 17, [37].

[71] McFarlane, in (n 2) 778, thinks that a duty imposed by proprietary estoppel creates a right in C's favour before any judicial order is made and LRA 2002 merely confirms this. This view is shared by Liew in (n 2) 224. In his secondary rights approach, C obtains 'an unliquidated secondary right when B breaches his or her primary duty, [which] is liquidated at the day of judgment'.

[72] Hayton, 'Constructive trusts of homes - a bold approach' (n 2) 486.

[73] Etherton LJ (n 21) 271.

[74] There is, however, an ambiguity in s 116(a) between a strong and weak meaning, as pointed out by Professor Simon Gardner in his lectures at Oxford, HT16. While it necessarily means that C has rights against D from the time of the required facts, on the strong meaning, C's rights will be thus fixed and be the same against X. On the weak meaning, the content of C's rights can vary from time to time, as the relevant facts change. As noted in Part 2,

Lord Walker's second reason is that proprietary estoppel requires only 'the minimum award necessary to do justice, which may sometimes lead to no more than a monetary award'.[75] In other words, whereas C would have deserved and received a proprietary remedy under CICT, assimilation with proprietary estoppel would result in C receiving only a monetary award. With respect, his argument is self-defeating: assuming *ex hypothesi* that the minimum remedy required to do justice is a proprietary remedy, making personal remedies available does not hinder the court's ability to award a proprietary remedy in any way. On the contrary, the greater remedial flexibility in proprietary estoppel vis-à-vis CICT is a policy reason for assimilation.[76] In the following discussion of the family proprietary estoppel, it will be shown that this remedial flexibility can be attained in a principled way.

However, the full assimilation thesis is hamstrung by the transparent rationality objection to proprietary estoppel's remedial flexibility, as noted in Part 2.[77] Whereas other commentators simply glide over this difficulty,[78] Hayton, one of the earliest proponent of this thesis, claims that the underlying justification for both the CICT and proprietary estoppel is 'the discretionary prevention of unconscionable conduct'.[79] He anticipates the transparent rationality objection to the open-texturedness of 'unconscionability' by asserting that '[a]s time goes on, there should be an interpretative community of judges and legal advisers who can come to a significant consensus on what is unconscionable in particular circumstances in the light of decided cases like the already existing cases involving constructive trusts and proprietary estoppels.'[80]

Hayton's proposal surely underestimates humanity's propensity to disagree. As noted above, proprietary estoppel is a heterogeneous doctrine and the wide range of contexts in which it applies calls for an equally wide range of value judgments. Full assimilation would obscure CICT's uniquely compelling domestic context, which is vastly different from the considerations in cases involving unilateral mistake, common expectations or imperfect promise in a commercial context. 'Unconscionability' would be the indiscriminate catch-all, giving judges an unacceptably large latitude in determining liability and the mode and quantum of remedies. Not only would this inject an intolerable degree of uncertainty into property rights, it is illegitimate for the judiciary to wrest this power for itself.[81] But if we substitute unconscionability for a less abstract concept like detrimental reliance as an organising principle, then proprietary estoppel loses the remedial flexibility that typifies its allure in the first place.[82] In light of this pessimistic conclusion, we must look elsewhere for a solution.

B. Family Proprietary Estoppel: A New Proposal

As noted at the end of Part 1, the doctrine of CICT is unable to fully vindicate the rationale of the

---

the weak meaning is the better characterization of CICT and, by extension, the family proprietary estoppel. But it must be noted that this is an outstanding controversy that has not been resolved by the courts.

[75] As Gardner pointed out in fn 65 of 'Family Property Today' (n 21), speaking of 'the minimum necessary to do justice' merely begs the question what that is.

[76] This was noted by both McFarlane, in (n 2) 778, and by Liew in (n 2) 223, claiming that this would encourage 'a more efficacious legal system'.

[77] Gardner makes a similar point in 'Family Property Today' (n 21) 434: '[Arguing for full assimilation] would be nonetheless helpful if it were clear what does drive the [proprietary estoppel]: but that is notoriously not the case'.

[78] McFarlane simply calls proprietary estoppel 'a recognised means by which [D] can come under a duty to [C]', whereas Liew notes that it is 'the only equitable doctrine which enforces B's secondary duty to allow A informally to acquire proprietary rights', without pausing to wonder whether that uniqueness is justified.

[79] 'Equitable rights of cohabitees' (n 2) 380.

[80] ibid 385.

[81] The family proprietary thesis, as we will see, is better insulated to this objection.

[82] The measure of the relief would have to be tailored to correct for C's detrimental reliance, for instance.

Gardner-Davidson theory. Building on this theory, it is the vision of the family proprietary estoppel thesis to do so fully and in a principled manner. For the judiciary to fulfil this vision, the premise that the judges have a duty to develop the law in pursuit of a more coherent conception of justice is needed. Both foundations are contestable.

The strongest argument against the family proprietary estoppel is that the context-based dichotomy between domestic and commercial cases is unstable and unpredictable. Unsurprisingly, proponents of the full assimilation thesis[83] and legislative reform[84] endorse this view. However, it is submitted that the very fact that the lower courts are dealing with CICT cases containing both familial and commercial elements, ostensibly without any grave difficulties ought to diminish the force of these criticisms.[85] After all, law operates by imposing categories onto an unruly world. It is trite wisdom that legal categories, so analytically elegant and immensely sensible at the time of conception, have a habit of being reduced to absurdity by the vicissitudes of life. So long as the dichotomy works most of the time, when borderline cases arises, all we can ask for is the judge to pay meticulous attention to factors such as the sophistication of parties and the relative significance of the commercial/domestic elements of their relationship, with the understanding that the 'domestic context' justifies both the law's remedial flexibility and circumvention of formality requirements relating to land.[86] Furthermore, it is apt to point to similar dichotomies in other areas of English private law, namely one that is firmly established in contract law[87] and another that is emerging in express trusts.[88]

With that out of the way, we can examine the main argument for judicial development of the family proprietary estoppel, i.e. a proper understanding of the normative positions of C and D requires extending a greater range of remedies to cases currently caught under CICT and abandoning some of the doctrinal limitations imposed by CICT and proprietary estoppel.

Firstly, it is central to the Gardner-Davidson theory that the law should treat cohabitation and

---

[83] McFarlane in (n 2) 776–7, calls it 'an unprincipled and unstable distinction' and that the domestic context is merely 'a *factual* difference to the *factual* question of whether [D] can be found to have made an implied commitment to give [C] a share of the benefit in [D]'s Freehold or Lease'. Liew, in (n 2) 217, argues that 'the domestic/business distinction is inherently unhelpful, causing uncertainty where the parties' use of the property or the parties' relationship is atypical'.

[84] Sloan (n 50) 126. He argues for a statutory alternative to estoppel for the enforcement of inheritance representation cases in emulation of New Zealand's Law Reform (Testamentary Promises) Act 1949. If the domestic context is not simply the 'circumstances of the case', but instead 'effectively determines where the decision-making process begins', then 'the attachment of importance to [the domestic context] is much more vulnerable to objection'.

[85] See n 7 and accompanying text. There are cases relating to investment properties (see *Geary v Rankine* [2012] EWCA Civ 555, [2012] 2 F.L.R. 1409, [18]) and properties bought by business associates (see *Gallarotti* [2012] EWCA Civ 865, [2012] 2 F.L.R. 1231, [6]).

[86] It is accepted that in inheritance representation cases, there must be an additional justification for circumventing formalities relating to wills, which is more well-known. This justification must come from the facts: in *Thorner*, we know that Peter's destruction of the old will leaving the farm to David and failure to make a new one are for reasons unrelated to David, and this would, it is submitted, provide the justification.

[87] The inaptly named 'intent to create legal relations' doctrine operates to police, *inter alia*, the boundaries between family arrangements where promises are presumed to be legally unenforceable, and commercial situations where they are. See e.g. *Rose & Frank Co v JR Crompton & Bros Ltd* [1924] UKHL 2 and *Balfour v Balfour* [1919] 2 KB 571 (CA).

[88] First raised by Lord Browne-Wilkinson in *Target Holdings Ltd v Redferns* [1995] UKHL 10, [1996] AC 421, 436, this has been echoed in *AIB Group (UK) Plc v Mark Redler & Co* [2014] UKSC 58, [70]-[71] per Lord Toulson and [102] per Lord Reed. Their Lordships were quick to emphasize this distinction is not categorical, however.

marriage the same way.[89] It should be noted that many European countries have legislated as such.[90] The sociological findings are clear. It is very unlikely that people decide not to marry because of the legal consequences, firstly because few people even know this area of law in the first place,[91] and secondly because even those who do are more likely to base their decision to marry on religious and social views, or to be influenced by their families, friends, and culture.[92] This ought to allay any fears that extending rights to cohabiting couples would undermine the legal institution of marriage or the individual's autonomy. As such, it is no coincidence that 'fairness' under the Gardner-Davidson theory resembles how the courts apply the Matrimonial Causes Act 1973, ss 24 and 25, to divorcing couples.[93] But there is one aspect which is dissimilar: under the statutory regime for divorcing couples, C can claim a monetary award as well as a beneficial interest. For cohabiting couples, C is invariably awarded a proprietary interest under CICT. This has practical consequences. Consider a reverse-*Kernott* scenario: if the value of the house C happens to be living in is the cheaper of the two, even if C receives all the beneficial interest in the house under the CICT, C remains undercompensated according to the deeper conception of equality displayed in *Kernott*. If the law should treat cohabitation and marriage the same way, in this scenario we have a strong reason to escape the confines of a constructive trust and seek the greater remedial flexibility offered by the family proprietary estoppel, which can give C simultaneously all the beneficial interest in the house *and* a monetary award. This principle of consistent treatment also legitimizes *judicial* development of the family proprietary estoppel: the legislation already in place gives democratic imprimatur to judges who are duty-bound to ensure that the law treats like cases alike.[94]

Secondly, it is submitted that the NIDR of C and D ought to be based entirely on the nature of their relationship, with no reference to whether D has a freehold (or leasehold). In principle, C equally deserves a share of D's assets, even if D does not own a home and C and D are long-time contractual licensees. Under the CICT, effectuating the implications of their relationship is made conditional in this way, per (1) and (5'), and is restricted to proprietary relief. There ought to be no such limitations under the family proprietary estoppel, providing another reason for its remedial flexibility.[95] In fact, this need for this flexibility will only sharpen, in light of 'the increasing number of situations where the parties' home is rented rather than owner-occupied'.[96] Admittedly, without a high-value asset like a family home, any claim against D would be contingent on D holding other types of assets and might not be viable in

---

[89] Gardner and Davidson, 'The future of Stack v Dowden' (2011) 127 LQR 13, 18. As they concede, the Law Commission seems to disagree, in its consultation paper on this area. See Law Commission, 'Cohabitation: the Financial Consequences of Relationship Breakdown' (n 67) at paras 6.23, 6.48-6.49, 6.62-6.66, 6.92-6.114. But note that under the Inheritance (Provision for Family and Dependants) Act 1975, if a cohabiting couple's relationship dissolves due to D's death and D's will does not give adequate financial protection to C, then the court is empowered to give C financial support from D's estate, as if C and D are married. This is certainly a *legislative* acknowledgement their normative positions are at least identical in this respect.

[90] Thorpe LJ, 'Property rights on family breakdown', (2002) 32 Family Law 891, 893.

[91] C Smart and P Stevens, *Cohabitation Breakdown* (FPSC/Rowntree 2000).

[92] J Herring, *Family Law*, (5th edn, Longman Publishing 2011), 94.

[93] *White v White* [2000] UKHL 54, [2001] 1 AC 596, 599-600, authoritatively states that, assets should be divided equally unless the parties' have more specific needs, reflecting a more general moral obligation on each of them to share their resources created by the materially communal relationship that is marriage.

[94] See n 89.

[95] This might lead to a semantic worry: why call it family *proprietary* estoppel if there is no *real* property involved? Put this way, the worry dissolves: firstly, reference will inevitably be made to the property (tangible or intangible) of C and D, which in the modern age is increasingly unlikely to be held in land, and, secondly, it could be read as an estoppel of *family property*.

[96] According to the 'English Housing Survey, Headline Report 2013-14' conducted by the Department for Communities and Local Government, English homeownership rate was at 63%, down from 70.9% in 2003, a level last seen three decades ago. See, also, Gardner 'Problems in Family Property' (n 13) 312.

practice.

Thirdly, on a related note, it is submitted that (2') should be tweaked: any remedy should not be conditional upon whether C has any belief about D's land. Consider a twist in the facts of *Thorner*: suppose that David spent all his years working unpaid in Peter's *studio* on the basis that he will inherit Peter's *intellectual property rights* in a hit Christmas song. Think Hugh Grant in *About a Boy* (2002). If Peter fails to do so in his will, David should nonetheless be able to claim from Peter's estate.[97] The centrality of land to personal wealth is in many ways a contingent upon the structure of the national economy and has been steadily declining from its zenith in feudal society. Understanding the rationale of family proprietary estoppel in this manner allows us to see the requirement of C's belief about D's land (if indeed this requirement exists) as non-essential to effectuating NIDR.

Taking the arguments above as a whole, it is submitted that the family proprietary estoppel has strong foundations in existing legal principles. But more than that: suppose there is a judicial duty to develop the law to further a more coherent conception of justice. As shown above, the family proprietary estoppel is necessary for the law to treat normatively identical Cs and Ds the same way. If that is right, it is obligatory and not merely supererogatory for the judiciary to develop the family proprietary estoppel.

Having fully explored what is demanded by principle, there are two further instrumental advantages offered by the family proprietary estoppel.

Firstly, the remedial flexibility of a family proprietary estoppel might become potentially helpful where the relationship is materially non-communal. Accepting that the driver here is reversing C's unjust enrichment of D, it is not at all clear that giving C a proprietary interest under CICT, as opposed to a monetary award, accords with the logic of unjust enrichment.[98] Given this is 'one of the most difficult, and certainly unsettled, issues in the law of restitution',[99] it is submitted that a family proprietary estoppel is future-proof: if and when this issue is settled by restitution lawyers, the judges will have the option of giving C a monetary award, if that is so required, under a family proprietary estoppel, rather than the compulsion of granting a proprietary interest under a CICT.

Secondly, it is submitted that, by assimilating the inheritance representation cases within family proprietary estoppel, the feminist criticism of the Gardner-Davidson account of the CICT can be met.[100] Bottomley argued the Gardner-Davidson communitarian approach is too prescriptive of family relationships and runs the risk of determining legal rights on the basis of patriarchal norms, such as finding 'communality' in relationships with classic gender roles. In the intervening twenty years since this argument was made, its force has diminished significantly. Today, English law recognizes same-sex marriage and civil partnerships, whereas material communality in CICT (on the Gardner-Davidson theory) already applies across the board to all types of family relationships. The family proprietary estoppel would finally consign this criticism to history. Juxtaposing the material communality of cohabiting couples next to that of prospective heirs who spend much of their lives working without pay is

---

[97] Admittedly, this factual matrix is freakish – or merely prescient of things to come.

[98] Gardner, 'Family Property Today' (n 21) 439.

[99] As Gardner pointed out in fn 78 of ibid: 'According to "the powerful argument" advanced in Birks, *Unjust Enrichment* (OUP 2003), ch 8, unjust enrichment relief should be proprietary where the claim involves an initial failure of basis, as where C mistakenly believes that he already has an interest in the house; but personal where the claim involves a subsequent failure of basis, as where D fails to perform his promise to give C such an interest. In practice, applying this distinction would often be difficult in our situation, given the rather unfocused quality of the typical facts.'

[100] A Bottomley, 'Women and Trust(s): Portraying the Family in the Gallery of Law' in S Bright and J Dewar (eds), *Land Law. Themes and Perspectives* (OUP 1998).

a powerful reminder of the diverse forms of familial organization and that gender roles, blood ties, romance, sexual fidelity, and so on are non-essential to the creation of relationships with normative implications.

C. An Act of Parliament?

As stated in Part 1, the law in this area is in disarray and there have been calls for legislative reform of both CICT[101] and proprietary estoppel.[102] The government announced in 2011 that it would not take forward the Law Commission's recommendations for reform of CICT during the 2010-2015 parliamentary term,[103] presumably because doing so might be framed politically as undermining marriage. The weakness of this argument has been pointed out above. The Law Commission has no plans to reform proprietary estoppel. The fact that legislation is not forthcoming in both areas of law is itself fatal to proponents of legislative reform.

But if this essay is right, it means that parties caught under the family proprietary estoppel have certain proprietary entitlements under the law. Though these entitlements are subject to a degree of judicial discretion, under s 116(a) of the Land Registration Act 2002, they arise from the time of the facts as an inchoate 'equity'. As such, they cannot be taken away in a manner that constitutes a disproportionate interference with rights under European Convention of Human Rights First Protocol Article 1, which protects the peaceful enjoyment of possession, and Article 8, which ensures respect for one's private and family life and one's home.[104] This should militate against any rashly hatched legislative proposals.

4. *Conclusion*

To complain that the law on CICT is unsatisfactory has become clichéd. This essay has presented a principled case for the judiciary to unite the CICT with parts of proprietary estoppel, resulting in the birth of the family proprietary estoppel. This is the best proposal for legal reform. Full assimilation is overly ambitious and there is no feasible proposal for legislative reform.

Hayward's verdict of the legal developments culminating in *Kernott* is sound: '[O]wing to the well-documented failings of the trust framework by the judiciary, academics and reform bodies coupled with the fact that statutory cohabitation reform is not imminent', the judiciary's incremental development of the CICT is 'an avowedly imperfect yet truly necessary development'.[105] Against this backdrop, the family proprietary estoppel represents the next step in the familialization of property law. Equity is not past the age of child bearing indeed.

---

[101] Law Commission (n 67).
[102] Sloan (n 50).
[103] Law Commission on Cohabitation <http://www.lawcom.gov.uk/project/cohabitation/> accessed 23 March 2016.
[104] See e.g. Gardner, *An Introduction to Land Law* (n 39) 196-7 for a brief discussion of Convention rights and legislative reform vis-à-vis the Law Commission's proposals. For a more general discussion of the applicability of Convention rights to private property law, see A Goymour, 'Property and housing' in D Hoffman (ed), *The Impact of the Human Rights Act on Private Law* (CUP 2011).
[105] A Hayward, '"Family property" and the process of "familialisation" of property law' [2012] CFLQ 284, 303.

'Where there is discord, may we bring harmony': *AIB Group (UK) v Mark Redler* and the Perils Facing Equity

Matthew Hoyle*

Introduction

Though the quote in the title is often (wrongly) attributed to St Francis of Assisi, it has now become synonymous not with harmony but with the immense conflicts of the 1980s and a controversial legacy that persists to the present day. In *AIB Group (UK) plc v Mark Redler & Co Solicitors* ('***AIB***')[1] the Supreme Court has, in what their Lordships viewed as a common-sense attempt to harmonize basic elements of common law and equity, instead taken a step down a worrying and conflict-ridden path. As noted below, the law in this area was hardly dispute-free before the decision, but the judgments delivered in this case suggest that, on this matter, the Supreme Court will not be moved. This step will likely be locked into the law of trusts for a long time to come and its implications should not be understated.

This essay will argue that while the decision may represent the authoritative statement on the law for now, normatively it cannot be justified by the reasoning given by their Lordships when one takes account of the differences between the equitable obligations at play in this case and contractual principles that were applied. Equally, it will note the failure to observe those differences and the resulting erosion of the separation between common law and equity represent a significant attack on the very foundation of the law of trusts. While it is unlikely to be overruled in the immediate future, it should be confined as narrowly as possible.

1. *The law on wrongful disbursement before* AIB

The traditional remedy for breach of trust by wrongful disbursement was that the beneficiary could view the accounts and falsify any unauthorised transactions under the supervision of a Chancery Master. This meant the court would regard any such debits as nullities, and thus the actual value of the trust capital would then be below the amount that the ledger stated should be in the trust. The trustee would then have to restore the property *in specie* or make good the loss in money equivalent. There was no room for causative inquiry, even if that meant money that had been lost regardless of any breach was recoverable.[2] Over time, the viewing of the accounts and the falsification of the disbursement became mere fictions, as parties pursued their claims directly in court for the missing sum. However, the underlying

* St John's College, Oxford. I am grateful to Professor Luke Rostill for his support and for his comments on the earlier drafts of this essay. I would also like to thank Professor Joshua Getzler for providing his insight into the case law, and to the editors of the Journal for all their work on this essay. All errors of course remain my own.

[1] [2014] UKSC 58, [2015] AC 1503.

[2] For a particularly clear example, see *Cocker v Quayle* (1830) 1 Russ & M 535, 39 ER 206, which followed *Brice v Stokes* (1805) 11 Ves Jr 319, 32 ER 1111.

basis of the remedy was left untouched until the case of *Target Holdings Ltd v Redferns* (‘***Target***’).[3]

In *Target,* a lender had required its solicitors, Redferns, to acquire mortgage documentation before it lent money to a client. The solicitors failed to do so, but still paid out the loan – held on bare trust – at the insistence of the client. The client was later revealed to be part of an enrichment scheme of, at best, questionable legality, which caused substantial losses to the lender. The lender brought an action for breach of trust, claiming that the entire loan was paid in breach, and that accordingly they were entitled to have Redferns make up the full sum, most of which they would have lost irrespective of the breach. Lord Browne-Wilkinson, giving the only reasoned judgment, held that the liability of Redferns to provide equitable compensation was limited to the loss caused by their breach.[4] He also suggested that the commercial nature of the trust had affected the remedy available,[5] though not in a manner clear enough to form a distinctive doctrine of ‘commercial trusts’.

The legal analysis in this case is complicated by the fact that the mortgage documentation was acquired by the solicitors after the wrongful disbursement and so when the correct asset was received, the ‘value’ (used loosely) paid into the account fell into line with the amount initially disbursed – loss being assessed in equity at the date of trial rather than the date of breach. As a result, the only loss suffered was the fall in market value of the asset paid into the account, which was not caused by Redferns. In the opinion of several commentators, most notably the then Millett LJ in an extrajudicial publication, the amount Redferns was held liable to pay in equitable compensation by Lord Browne-Wilkinson was the same as it would have been on traditional accounting principles. The reasoning employed by Lord Browne-Wilkinson to reach that position was however significantly flawed[6] – Millett LJ also went further and argued that any outcome based on ‘compensation’ rather than enforcing performance was completely wrong.[7] Indeed, this issue comes back to haunt the decision in *AIB*. The controversy cast a shadow over the case and left the exact nature and operation of the law in real need of clarification or even an entire reworking. It is against this background that we arrive at the decision in *AIB,* the Supreme Court’s first chance to properly revisit the decision in *Target.*

2. *The facts of* AIB *and the case at first instance*

The Sondhi family wished to borrow £3.3m from AIB against a property then worth £4.25m. The bank agreed, provided that they could have first legal charge over the property. However, Barclays Bank held the first charge, worth £1.5m split over two accounts of £1.2m and £300,000. AIB therefore instructed the solicitors, Mark Redler & Co (acting for both sides) to hold the loan on bare trust, and to pay it out once the solicitors had redeemed the Barclays charge and established a first charge for AIB. Redler redeemed the £1.2m charge, but failed to redeem the £300,000 charge, leaving AIB without the requisite first charge. They then paid out the unapplied trust capital, with the £300,000 that should have been used to redeem the charge, to the Sondhis.

The value of the property subsequently collapsed and the Sondhis defaulted. Barclays then exercised their charge on the property, selling it for around £1.2m. AIB was left with the remainder once the £300,000 had been recovered, i.e. some £800,000. Finding themselves substantially out of pocket, they sued Redler, claiming that by failing to redeem both charges

[3] [1995] UKHL 10, [1996] AC 421.
[4] *Target* (n 3) 439, 440 (Lord Browne-Wilkinson).
[5] ibid 435-436.
[6] PJ Millett, ‘Equity’s Place in the Law of Commerce’ (1998) 114 LQR 214, 227.
[7] ibid 225. For further analysis see Paul Davies, ‘Remedies for Breach of Trust’ [2015] 78 MLR 681.

before paying the Sondhis, Redler had disbursed the capital in breach of trust. Redler would have to account for the trust capital, making up any shortfall. Once the approximately £800,000 recovered from the sale was taken into account, with benefits being offset against liability for breach, this meant that Redler would be liable for around £2.4m, approximately half of which would have been lost regardless due to the fall in property prices. Redler admitted liability for breach of contract and in negligence, but not breach of trust, and argued that they could only be liable to the contractual or tortious standard.

The trial judge held that while there was a breach of trust, only the £300,000 was wrongfully disbursed, awarding equitable compensation for that amount. Both parties appealed the judgment.

3. *The decision by the Court of Appeal*

The Court of Appeal reversed the trial judgment in part, holding that the entire sum was paid out in breach. They did not, however, allow AIB to claim the full sum, holding that the precedent of *Target Holdings Lt. v Redferns*[8] meant compensation for breach by wrongful disbursement was assessed on a causal basis.[9] Thus, where trust capital would have been lost even if it had been lawfully applied, a beneficiary can only recover compensation for the capital loss *caused* by the breach. As Redler had only caused the loss of the £300,000, that was the extent of their liability to AIB. The rest of the capital was lost due to a fall in house prices, and would have been lost even in the absence of breach. AIB appealed to the Supreme Court.

4. *The decision by the Supreme Court*

While the case was on appeal to the Supreme Court, it garnered significant academic attention. The attack on *Target* presented the Court with an opportunity to revisit, clarify and hopefully overrule a problematic decision.[10] The Court in *AIB* was faced with two theoretical questions. First, was Lord Browne-Wilkinson's reasoning in *Target*, and not just the outcome of which, correct, i.e. was he right to hold that equitable compensation for wrongful disbursement was to be determined on a causal basis? Secondly, did this norm apply only to commercial trusts, or did it apply to all trusts?

The hopes of those who opposed *Target* were not fulfilled however. The Court held that the Court of Appeal was correct, and that Lord Browne-Wilkinson's causation-based reasoning in *Target* was sound: AIB was only entitled to the £300,000 that Redler's breach had caused to be lost.[11] The Court also held that there were no substantial differences in the principles underpinning 'commercial' and 'traditional' trusts,[12] although Lord Toulson suggested that in cases of 'commercial' trusts the commercial and contractual nature and purpose of the transaction may 'have a bearing on the appropriate relief' relative to a case involving a 'traditional' trust.[13]

Despite their Lordships seeing the ruling as simply approving *Target* and the subsequent jurisprudence therefrom, it resolves a number of standing questions about remedies for breach of trust and the role of 'commercial' trusts and rescues *Target* from its

[8] *Target* (n 3).
[9] [2013] EWCA CIV 45, [2013] PNLR 19 [47]
[10] Davies (n 7) 683.
[11] *AIB* (n 1) [62]-[68] (Lord Toulson), [114] (Lord Reed).
[12] Ibid [70] (Lord Toulson): 'Lord Browne-Wilkinson did not suggest that the principles of equity differ according to the nature of the trust...'
[13] ibid [70]-[71] (Lord Toulson), [102] (Lord Reed).

legal purgatory.[14] Indeed, while *Target* may have laid the groundwork for the new legal paradigm, it lacked clarity and was mired in controversy. Whatever one thinks of it merits, in resolutely defending and comprehensively restating the new basis of recovery, *AIB* represents a significant case which will play a central role in the law of equitable compensation and remedies for breach of trust in the years to come.

5. *The Supreme Court's reasoning*

A. Precedent

Their Lordships appeared to assume that the foundations on which they decided the case were now thoroughly established in the law.[15] But this arguably overlooks a substantial quantity of academic commentary questioning the soundness of the reasoning in *Target* and a line of authorities going back to 1801 providing for a very distinct and separate basis of recovery to that expounded in *Target* and now *AIB*.[16] The vast weight of legal history goes against the approach taken in these cases.

Of course, an argument from tradition alone is not enough to justify taking a 'backwards step'[17] to a previous rule. But the case law relied upon is not limited to Georgian precedents. For example, shortly before *AIB* was decided, Lord Sumption used the assumption that accounting was the enforcement of a primary and not a secondary obligation to underpin the distinction between 'types' of constructive trustee.[18] It is not merely a matter of defunct tradition but a well embedded part of the framework of English equity. This is perhaps why besides *Target*, their Lordships were forced to look to precedents from other common law jurisdictions.

B. Precedents from other common law jurisdictions

Much emphasis was placed by their Lordships on the development of equitable compensation in other jurisdictions,[19] particularly McLachlin J's judgment in *Canson Enterprises Ltd v Boughton & Co*,[20] ('***Canson***') which was also a key influence on Lord Browne-Wilkinson in *Target.*

Arguably, *Canson* was not the strongest precedent for their Lordships to rely on at all, much less the extent that they did. As noted in *Hayton and Mitchell on the Law of Trusts & Equitable Remedies*,[21] it did not concern a trust relationship, merely a fiduciary one. As will be argued below, the traditional remedy is of paramount importance in a trust relationship, given the nature of a trust, and the same is not necessarily true for other company or contract-based applications of fiduciary law. Even McLachlin J conceded that 'equity is concerned, not only to compensate the plaintiff, but to enforce the trust which is at its heart',[22] and this principle would seem to go against the conclusions their Lordships in both *Target* and *AIB* drew from the case.

---

[14] Davies (n 7) 681, noting even Lord Toulson (*AIB* (n 1) [20]) acknowledged *Target* to be largely unloved.
[15] *AIB* (n 1) [63] (Lord Toulson).
[16] ibid [47] (Lord Toulson).
[17] ibid [63] (Lord Toulson).
[18] *Williams v Central Bank of Nigeria* [2014] UKSC 10, [2014] 2 WLR 355, [13].
[19] *AIB* (n 1) [66] (Lord Toulson).
[20] [1991] 3 SCR 534 (Supreme Court of Canada).
[21] Ben McFarlane and Charles Mitchell, *Hayton and Mitchell on the Law of Trusts and Equitable Remedies: Text, Cases & Materials* (14th edn, Sweet & Maxwell 2015), 10-057.
[22] *Canson* (n 20) [61].

If that was the principle she was aiming to apply, the result in McLachlin J's arguably misunderstood the dichotomy between substitutive and reparative compensation.[23] Per the analysis by Elliot and Mitchell[24] the latter is a secondary remedy, repairing the pecuniary loss suffered by the claimant due to the breach of primary duty. The former however is a primary remedy, which provides a monetary substitute for the performance of the duty, in this case the trustee's stewardship obligation. As Elliot and Edelman noted,[25] McLachlin J's reasoning confusingly mixes together ideas of substitutive compensation and restitution of trust property as if what is being substituted is the reconstitution of the fund, not the performance of the stewardship obligation.[26] Given performance of the latter is not causal based, a case which does not recognise this characteristic of the substitutive remedy does not provide a solid foundation for a court to declare that compensation must be causal based. It is perhaps this misunderstanding spreading from *Canson* into English law through *Target* that led Lord Toulson in *AIB* to declare that the aim of reparative and substitutive compensation is the same: to 'make good any loss suffered'.[27] To *repair* losses is the aim of reparative compensation. However, as noted above, substitutive compensation is a *substitute* for the performance of the primary obligation, i.e. to not wrongfully disburse the trust capital. The breach is in the disbursement and not in the loss suffered, and so the trustee must provide a money substitute for the level of disbursement. If this is not the distinction between the two, then it is hard to understand what distinction there is beyond their name, even if logically they cannot be the same.

As a result, it is highly questionable to suggest that a weak precedent from another common law jurisdiction is sufficient to displace the numerous historical and current precedents detailing the accounting process for falsification.

C. 'Fairy tales'

Moving beyond precedent in his search for support for his view that the traditional accounting principle ought to be abandoned, Lord Toulson set his sights on the legal fictions present in accounting. It is true that there is no longer any examination of the trust ledger by a Chancery Master, where the beneficiary scores through and thus falsifies disbursements. This to him was unacceptable: the law was 'wrong' when it relied on 'fairy tales'.[28] He may have a point insofar as there is no reason to preserve legal fictions for their own sake. However, where legal fictions were previously present, the solution has usually been to get rid of them while maintaining the underlying mechanism,[29] not to change the whole mechanism, and throw the metaphorical baby out with the bathwater. The presence of a legal fiction alone, where it does not make the law practically unworkable or difficult, is not sufficient grounds to dispense with the underlying mechanism, especially if it is well established in the law, as is the case here.

---

[23] McFarlane and Mitchell (n 21) 10-057.
[24] Steven B Elliot and Charles Mitchell, 'Remedies for Dishonest Assistance' (2004) 67 MLR 16, 23-25.
[25] Charles Mitchell, 'Equitable Compensation for Breach of Fiduciary Duty' (2013) 66 CLP 307, 19-20.
[26] For further analysis, see McFarlane and Mitchell (n 21) 10-057, 13-049.
[27] *AIB* (n 1) [54] and [66]-[67] (Lord Toulson).
[28] ibid [69] (Lord Toulson).
[29] See, for example, the action for Detinue, which was based on losing and finding of property, even though they became mere fictions long before the action was abolished: W.E. Peel and J. Goudkamp, *Winfield and Jolowicz on Tort* (19th edn, Sweet and Maxwell) at 18-006. Further, see unjust enrichment replacing 'quasi-contract': Peter Birks, *The Roman Law of Obligations* (Ed. Eric Descheemaeker, 1st edn, OUP 2014) 258-9; or 'implied' or 'constructive' contract: Peter Birks, *Unjust Enrichment* (2nd edn, OUP 2015) 4, 267-270, 271.

Something more is needed to justify the mechanism's dispensation than an assertion that it is too archaic. That the existence of the mechanism is the reason for a substantively unfair result might be a good reason. Section D, however, argues that there is no substantive unfairness.

Alternatively, any assertion that its form obscures the law, as Lord Reed did,[30] is unconvincing. The accounting mechanism is simple and workable. It is a disservice to lay parties to suggest they cannot understand that if money is wrongly paid out of a fund, the law treats the money as if it is still in the fund until the asset that was meant to be acquired is indeed acquired. If that asset is never acquired, then the amount 'in' the fund will not match the amount noted in the ledger and therefore the person who has acted wrongly in making that disbursal should correct the perceived shortfall. It is no more mysterious or confusing than the law treating certain contracts[31] or ultra vires actions which clearly do, in reality, exist as void *ab initio*, or courts closing their eyes to inadmissible evidence they are nevertheless actually aware of.[32] Though Lord Reed[33] and Ho[34] may suggest that this was a victory for transparency, it is hard to see why a substantive rather than a mere formal change was necessary in this case.

D. Relationship between loss and compensation

Both Lords Reed[35] and Toulson[36] argued that compensation for wrongful disbursement should be calculated from the loss caused, as it is in contract and tort. This is wrong on two grounds. First, it presumes that wrongful disbursement generates a secondary duty to pay compensation, and secondly, it presumes that compensation must be linked directly to loss caused.

The first problem arises when one treats the breach of the primary obligation to steward the trust fund (in this case by wrongful disbursement) as leading to a secondary obligation to compensate, as if it were analogous to a contractual undertaking to look after the capital, or a bailment relationship.[37]

As noted by then Millett LJ[38] shortly after the *Target* decision, that in cases such as this, the court is merely calling upon the trustee to perform his duty and account for the trust capital. The trustee is not burdened with any additional *secondary* obligation; the compensation he pays is in performance of his *primary* duty. This can be contrasted with the secondary obligation arising from breach of the duty to exercise reasonable skill and care in a case of wilful default, which entails fulfilling a secondary duty to add into the account the sum that should have been made in profit.

On this analysis, it does not matter how the property *in specie* or its money equivalent ends up in the fund, provided all that should be there actually is. Here, a causal analysis of compensation makes little sense; the trustee is not paying compensation for breach of an

[30] *AIB* (n 1) [137]-[138] (Lord Reed).
[31] See for example those void for fundamental mistake in *Bell v Lever Brothers Ltd* [1931] UKHL 2, [1932] AC 161.
[32] See for example the court ignoring illegal evidence in *R(Bancoult)* [2008] UKHL 61 or the effect of evidential presumptions on illegality in *Tinsley v Milligan* [1993] UKHL 3, [1993] 3 WLR 126.
[33] *AIB* (n 1) [138] (Lord Reed).
[34] Lusina Ho, 'Equitable Compensation on the Road to Damascus?' (2015) 131 LQR 213, 215.
[35] *AIB* (n 1) [62]-[63] (Lord Toulson), [93] (Lord Reed).
[36] ibid [136] (Lord Reed).
[37] ibid [93] (Lord Reed).
[38] *Bristol & West Building Society v Mothew* [1996] EWCA Civ 533, [1998] Ch 1; Millett (n 6) 224-227.

obligation, primary or secondary. He is simply accounting for the amount that should have been there if he had not disbursed capital in breach of trust. Therefore, Redler should have produced, upon request, the full amount minus offsets, i.e. some £1.3m.[39]

Even if one disagrees with this analysis and believes there is breach that should entail 'equitable' compensation, why must we then adopt contractual principles of recovery that state that breach of the obligation to perform leads to an obligation to make good loss caused? The answer is that we need (and indeed must) not. The reason lies in the differing interests of the parties and natures of the relationship in trusts and contract cases, which will be dealt with in Section E.

This leads to the second problem: that the Court's notion of compensation in *AIB* is not the only type of equitable compensation. As noted above, reparative compensation is only one form of compensation, and often may not protect the beneficiaries' interest adequately. As the Supreme Court has observed in *Cavendish Square Holdings BV v Talal El Makdessi* ('***Cavendish Square***'), '[reparative] compensation is not necessarily the only legitimate interest that the innocent party may have in the performance of the defaulter's primary obligations'.[40] Indeed, the nature of the primary obligation in a trust relationship has lead equity to prefer a primary remedy: an enforcement of the stewardship obligation. Compensation of this kind is substitutive compensation[41] (which is not, by its nature, causal based) – as a substitute for performance – and this is consistent with the traditional remedy provided by falsification. As recognised by Millett LJ in *Bristol v Mothew*,[42] when the trustee has breached his primary stewardship obligation, he is compelled by the court to provide a substitute for the capital lost due to his wrongful disbursement. The quantum of that substitute is the difference between the capital actually in the fund and the value that should be in the fund, assessed from the time of trial. There is no need to engage in a causal inquiry to set the level of compensation.

Perhaps as the court had determined that compensation, whatever name it went by, should be reparative,[43] neither of these arguments were accepted by the court in *AIB*, but they are arguably superior in terms of conformity to the principles of equity outlined below and are more consistent with precedent on substitutive compensation.

Even if one does reject the court's analysis of why the traditional remedy is inappropriate, there can be no doubt that liability for breach of trust, on such a basis, is swingeing. Lord Toulson, however, went a step further and labelled it 'penal'.[44] While it may not be a specific reference to the contractual doctrine, it taps into the same principle as that in contracts: payment of a disproportionate sum for breach of an obligation, one unrelated to loss incurred by the breach.[45] However, this analysis must be rejected.

Enforcement of a primary obligation rather than a secondary obligation may often lead to liability in excess of loss 'caused'.[46] The Supreme Court noted in *Cavendish Square* that a primary obligation and thus its direct enforcement, no matter how burdensome, cannot be

---

[39] This being the amount disbursed (£3.3m) minus the £1.2m rightfully applied, and the £800,000 recovered from the exercise of the charge.
[40] [2015] UKSC 67, [2015] 3 WLR 1373 at [32] (Lords Neuberger and Sumption), though in the context of contract.
[41] Elliot and Mitchell (n 24) 23-25.
[42] *Mothew* n 38.
[43] *AIB* (n 1) [64] and [70] (Lord Toulson).
[44] ibid [64] (Lord Toulson).
[45] *Dunlop Pneumatic Tyre Co Ltd v New Garage & Motor Co Ltd* [1914] UKHL 1, [1915] AC 79; *Cavendish Square Holdings BV* (n 40).
[46] See the discussion of *White & Carter (Councils) Ltd v McGregor* [1961] UKHL 5, [1962] AC 413, especially Lord Reid at 431.

'penal' at common law.[47] Indeed, the contractual remedies enforcing primary obligations – specific performance and the action for the price – may entail a financial burden on the defendant well in excess of the loss he 'caused'. The action for the price may, as Ho argued,[48] be a pre-agreed sum, but so is a penalty clause. Its validity is not derived from the time when it is agreed, but on what the liability enforces. One cannot, by tacit reference to the principles underlying contract law, judge the accounting remedy penal, without setting oneself against a significant body of that same law. Suffice to say, something more than establishing that the traditional principles of recovery *may* seem harsh upon trustees (though not even necessarily, by common law standards) is needed to establish why those principles ought to be abandoned. This is especially true if equity is pursuing a different objective to contract law, which is a matter that their Lordships did not sufficiently consider.

E. Misunderstanding of the nature of and policy underlying equitable relationships

Their Lordships in this case treat one set of principles of recovery as applying generally across common law and equity and, in doing so, have begun to generalize the principles behind very distinct legal obligations.[49] In aligning the general principles of recovery at common law and in equity,[50] they are blurring the boundary between an equitable relationship and a contractual one, each a very different beast. As Lord Millett noted,[51] trusts law is not the mirror image of contract that simply arose in the equitable jurisdiction for historical reasons. The two are underpinned by very different legal policies, and this should not be forgotten. In scrapping the traditional remedy and moving towards a model of compensation roughly analogous to contract, the Court arguably has made a significant and worrying misstep.

The English law of contract (setting aside covenants) is predicated on the notion of a 'bargain', where parties of equal bargaining power act for their own financial interest. There is no general notion of 'good faith' in our law of contract,[52] and any suggestion there should be has been repeatedly rejected – parties should not be required to look after each other's interests. In contrast to this, trust relationships are not predicated on consideration, even those arising from contract. The relationship between the parties is one of a fiduciary nature, where one party must act in the others best interest. He has 'subordinated his own interests to those of the beneficiary',[53] and this subordination has consequences.

The primary obligations attaching to contracts are self-interested and bilateral; the principle of quid pro quo underpins the relationship. The parties definitively do not have a specific right to what they have contracted for,[54] and arguably freedom of contract demands that parties be free to breach contracts provided they bear the loss caused.[55] The primary obligation in a trust on the other hand is one of stewardship: to look after the beneficiary's

---

[47] *Cavendish Square* (n 40) [13] (Lords Neuberger and Sumption).
[48] Ho (n 34) 217-218.
[49] *AIB* (n 1) [93], [136] (Lord Reed).
[50] ibid [64] (Lord Toulson): 'all agree' that 'the basic purpose of *any remedy* (my emphasis) will be either to put the beneficiary in the same position as if the breach had not occurred (…) Placing the beneficiary in the same position as he would have been in but for the breach may involve restoring the value of something lost by the breach or making good financial damage caused by the breach.' This is the accepted basic common law principles recovery, treated as a general principle.
[51] Millett (n 6) 225.
[52] General duties of good faith, being contrary to the nature of contract, cannot even be contracted for: see *Walford v Miles* [1992] 2 AC 128 (HL).
[53] *Novoship (UK) Ltd v Mikhaylyuk* [2014] EWCA Civ 908, [2015] QB 499, [68].
[54] At least not at common law, though certain specifically enforceable contracts may create property rights through a constructive trust. Even here, however, the right does not arise directly from the contractual obligation but the imposition of a trust.
[55] *Printing and Numerical Registering Co v Sampson* (1875) 19 Eq 462.

property[56] (or at least what has to be conveyed to them),[57] and remuneration is an additional rather than automatic matter.[58] By willingly entering into a trust relationship rather than or in addition to a simply contractual one, the trustee is willingly accepting a much more onerous task, and thus must accept that relationship's more burdensome duties and more wide-ranging liability.[59]

As a result, breach of contract and breach of trust demand different remedies. Preserving the party's freedom to act in their self interest in contract means they should usually only be liable in so far as they cause loss to the other party[60] or agree to a debt[61] or a fair sum.[62] There are no exemplary damages for breach of contract.[63] To award such damages would be contrary to the established policy that breach of contract alone without tortious conduct is not wrongdoing deserving of punitive response.[64] Further, it would be contrary to commercial policy to make parties liable for capital that would be lost anyway, as this is simply a reality of the business world that is the primary domain of contract, and contractual rules on damages generally reflect this. Such a loss would, for example, be caught out by the contractual rules of remoteness and is not usually within the reasonable contemplation of commercial parties.[65]

On the other hand, a trustee breaching his primary obligation has wrongly disposed of property to which other parties have absolute entitlement. He has failed in his responsibility to safeguard the interests of persons who, by virtue of being beneficiaries, are in a uniquely vulnerable position to him, and to whom he owes a duty of 'undivided loyalty'.[66] The legal policy articulated by the courts is one requiring 'maintenance of a very high standard of conduct'[67] by the trustee, and this has been consistently reaffirmed.[68] Indeed, such heightened standards are entirely orthodox in regard to trustees as fiduciaries. To protect principals, fiduciaries are subject to extremely restrictive anti-conflict duties which are intended to prevent fiduciaries from even considering placing themselves in potential conflicts.[69] As Lord Dunedin once remarked, the equitable jurisdiction will be used 'to keep persons in a fiduciary capacity up to their duty'.[70]

The same is true of trustees in regard to wrongful disbursement. To sufficiently protect beneficiaries, the courts have laid down a duty that discourages negligence or ambivalence to such risks by trustees and thus attempting to prevent the loss itself from arising. As a result, the remedy given by the court should be to make trustees perform their central, primary obligation, or else this duty and possibly the entire project of the law of trusts begins to come

---

[56] *Baker v Archer-Shee* [1927] UKHL 1, [1927] AC 844*; Foskett v McKeown* [2000] UKHL 29, [2001] 1 AC 102.
[57] *Saunders v Vautier* (1841) 4 Beav 115.
[58] *Keech v Sandford* (1726) Sel Cas 1 King 61*; Re Duke of Norfolk's Trust* [1982] Ch 61; see also the Trustee Act 2000, ss 28-30.
[59] Jamie Glister and James Lee (eds), *Hanbury & Martin: Modern Equity* (20th edn, Sweet & Maxwell 2015), 475, 18-001.
[60] *Robinson v Harman* (1848) 1 Ex 850.
[61] *White & Carter v McGregor* (n 46).
[62] *Dunlop* (n 45); *Cavendish Square Holdings* (n 40).
[63] *Addis v Gramophone Co Ltd* [1909] UKHL 1, [1909] AC 488.
[64] ibid 494, 496 (Lord Atkinson).
[65] See *Hadley v Baxendale* (1854) 9 Ex 341.
[66] *Mothew* (n 38), 18 (Millet LJ).
[67] *Novoship* (n 53) [76] (Longmore LJ) quoting in part Gibbs J in *Consul Development Pty Ltd v DPC Estates Pty Ltd* (1975) 132 CLR 373, 397.
[68] *FHR European Ventures v Cedar Capital* [2014] UKSC 45, [2015] 1 AC 250, [33].
[69] *Keech v Sandford* (n 58); *Boardman v Phipps* [1966] UKHL 2, [1967] 2 AC 46.
[70] [1914] AC 932 (HL), 963.

apart. Indeed, the Court of Appeal has recognized[71] that in the reverse situation (when utilising s 61 of the Trustee Act 1925 to relieve liability), when determining which actions leading to the breach could be assessed for unreasonableness, direct causation is not an appropriate test. Instead, in line with the underlying policy of trust law being the protection of the beneficiary, the courts have taken a liberal approach to the connection required between the wrongful conduct and loss complained.[72]

An even more expansive approach was outlined in *Cavendish Square*: the Supreme Court noted that in certain contexts, even purely contractual ones, the innocent party definitively possesses 'a legitimate interest in performance extending beyond the prospect of pecuniary compensation flowing directly from the breach in question'.[73] There is no reason why the general analysis does not apply even more pressingly in trust cases, and it is submitted in light of what has been argued that the nature of the relationship between trustee and beneficiary is so different from that between mere contracting parties that the legitimate interest of the beneficiary in enforcing the primary obligation far exceeds that of a contracting party. If commercial parties do not want such a relationship, they may remain in a purely contractual one, but they should not expect the rules protecting beneficiaries to be scaled back to suit them if they have chosen a trust relationship. Therefore, in answer to the question posed earlier,[74] the vast dichotomy between contractual and trust relationships and the much greater legitimacy of the beneficiary's interest in performance of the primary obligation of stewardship is why we should not follow the contractual model of making good only loss directly caused.

Even if one does not accept that the court should enforce the primary obligation through the remedy given, and instead should respond to secondary obligations generated by breach by awarding equitable compensation, it does not necessarily follow that that compensation must be linked to loss 'caused' in a strict sense. As noted above, other areas of the law of trusts, such as that in *Santander v RA*,[75] understand causation to extend much further than is suggested in *AIB*. The heightened standard of the primary obligation and the more serious potential consequences of its breach require us to widen the scope of the secondary obligation (liability in damages) to include not only losses directly 'caused' but also to the monies wrongfully disbursed in breach of trust. These are, after all, monies which the beneficiary has a paramount interest in protecting and having correctly administered, and an interest which goes much further than simply recovering only money that would not have been lost anyway.

To suggest that basic contractual principles be followed over basic equitable ones is to treat the trust relationship as something very different to the creature devised over hundreds of years by the chancery courts. It is not a mere contractual obligation to look after money, which would lend itself to rules of causation and remoteness. The blurring of the lines here clashes with the legal policy underpinning the remedies for breach, and this may cause issues across the law of trusts if it is carried any further.

---

[71] *Santander UK v RA Legal Solicitors* [2014] EWCA Civ 183, [2014] PNLR 20, Briggs LJ at [29]: 'I would, finally, *caution against an over-mechanistic application of the requirement to show the necessary connection between the conduct complained of and the lender's loss.* There may be highly unreasonable conduct which lies at the fringe of materiality in terms of causation, and only slightly unreasonable conduct which goes to the heart of a causation analysis. It would be wrong in my view to allow this purely mechanistic application of a causation-based test for the identification of relevant conduct to exclude the former from any consideration under s 61.' (emphasis added)

[72] ibid.

[73] *Cavendish Square Holdings* (n 40) [28].

[74] See the question posed at the beginning of Part 5, Section D, above.

[75] *Santander* (n 71).

Perhaps the most damning implication of this false equivalency in *AIB* is that a party can breach a trust (committing a wrong) without consequence, provided he does not cause loss. Of course, in many cases there will be loss 'caused' and thus some recovery, but cases like *Target* show this is not always the case - the actual loss to the beneficiary in cases like *Target* and *AIB* is always larger than the loss 'caused'. Indeed, this (often large) discrepancy is the very basis of the dispute. The beneficiary, vulnerable to the trustee and often reliant solely upon equity to ensure that his instructions are followed to the letter, must bear the loss when he has likely committed no wrong at all. A breach of trust not 'causing' loss should not be treated like a breach of contract causing no loss. Doing so fundamentally undermines the trustee-beneficiary fiduciary relationship.

Lord Reed even noted that different remedies are available because of the differing relationships in law and equity,[76] not merely because of historical coincidence: remedies should 'reflect the characteristics of the obligation'.[77] This supports having separate remedies. He then proceeds to say, however, that 'transparency' demands that where the principles are the same across equity and common law, the remedies should be made consistent, and where the principles are different, they should be kept different. For him, the underlying principles of equitable compensation in commercial trust cases are the same as those of common law damages, and so they should be harmonised and applied consistently. However, this reasoning is predicated on the assumption that the principles are the same in these cases, which, it has been noted, they are not. It fails to justify the assault on the nature of the trust obligation present in this case.

E. Conclusions on the Court's reasoning

This leaves little to be said for the Court's justifications. Of course, now that the change has been committed to, asserting the something different should have been decided is arguably not sufficient argument for the law to do a volte-face, in light of the importance of stability in the law. However, it is submitted that the very troubling implications for equity in the reasoning underpinning this decision and these potential detriments are a positive argument against standing by the now established post-*Target* approach.

6. *What should have been decided?*

A major problem for those who disapprove of *AIB* is that, while there are numerous criticisms to be made of their Lordships' judgments, those critics lack a coherent view of what the traditional remedy demands in this case. Whatever one may make of the court's reasoning, if the old model does not provide a useful answer, then simply as a matter of pragmatism we must accept the new one. This argument will be addressed below.

A. Possible alternative analyses

There appears to be a number of alternatives presented. As with *Target,* Lord Millett has (extrajudicially) suggested a means by which the substantive outcome of *AIB* may be correct even if the reasoning, in his opinion, is not.[78] He argues that though the disbursal of £3.3 million was in breach of trust, and therefore had to be restored by Redler, they could be discharged from this duty by producing an 'authorised substitute'. In this case, that would be the second charge which was acquired, less the amount by which it was subordinated to

[76] *AIB* (n 1) [92]-[93], [138].
[77] ibid [138] (Lord Reed).
[78] Lord Millett, 'Reflections on the Decision of the Supreme Court in *AIB v Redler',* lecture to the Professional Negligence Bar Association, 27 Jan 2015. Noted in McFarlane and Mitchell (n 21) 10-100.

Barclay's first charge. Redler would therefore only have to pay £300,000 into AIB's trust fund in order to have brought the fund into line with the trust as it should have been without the firm's negligence. Thus, even a substitutive performance claim by AIB would still only yield the £300,000 found to be the extent of liability in this case.

This claim has been met with scepticism,[79] particularly in regard to the use of the 'authorised substitute'. The Court of Appeal did not believe it to be an authorized substitute, and this finding was not appealed in the Supreme Court. It is difficult to see how in any sense a second charge could be an authorised substitute in this case, as the terms of the trust explicitly required a first charge be acquired. Thus, adopting this position would still weaken the beneficiary's position and control of a trust to an extent inconsistent with the principles outlined above.

More fundamentally, there is arguably little to gain by finding such solutions, as intellectually elegant as they are. If one believes their Lordships reasoned incorrectly, there is nothing objectionable in arguing their reasoning led to the wrong outcome, and equally there is little benefit to one's argument against the reasoning simply in saying it gets to the same result irrespective of the reasoning. Indeed, if all cases like *Target* and *AIB* can be given the same treatment, then from a purely practical standpoint it is difficult to justify why anyone should be so concerned with the exact reasoning if it has no bearing on the outcome. It would therefore be best to start from first principles in this case and follow those to the conclusion they entail, not twist those principles to return the 'right' number.

Given that AIB did not challenge the Court of Appeal's finding that the whole £3.3 million was wrongfully disbursed, their Lordships were asked to decide the extent of Redler's liability for that disbursement. The charge acquired by the solicitors, as it was a second charge, was of little value as an asset in itself as it was not what *AIB* asked them to acquire and therefore their acquisition of such a charge instead of a first charge is itself a breach of trust. However, the proceeds recovered by *AIB* from the sale itself, after Barclay's had recovered its first charge, can be offset. Thus, assessing losses from the date of trial, the amount wrongfully disbursed minus the amount recovered is £3.3 million minus the £800,000 leaving outstanding a total of around £2.5 million, which Redler was under a duty to restore to the trust fund.

B. The role of substitute assets

The Court of Appeal's finding that the £1.2 million spent recovering part of the charge was still a wrongful disbursement is, in the author's opinion, right, as there is a world of difference between a first charge and a second charge, as demonstrated in this case. Lord Millett's analysis raises a question in a more difficult context: where the incorrectly acquired asset is materially and economically identical to the one required, if the court is compelling the trustee to perform his primary duty, then is the trustee to be compelled to purchase the actual asset itself and dispose of the wrongly acquired asset? In the author's opinion he would not be required to do so, though that may be practically inevitable.

There is obviously a breach of trust the moment the trust capital is disbursed to acquire the wrong asset. Equity, however, unlike common law, assesses loss at the date of trial.[80] As a result, a great deal will turn on whether, in the interim period, the trustee has performed his primary duty and acquired the correct asset.

[79] McFarlane and Mitchell (n 21) 10-101.
[80] *Canson* (n 20) 554-555.

If he has, the court will acknowledge the receipt of the asset now in the fund[81] and (if it has not depreciated in value in the interim) there would be nothing for the court to compel and no possibility of substitutive compensation; the trial would be moot. In this scenario, it matters not whether the trustee sold the incorrectly acquired asset or used new monies in the purchase as the primary duty has been performed.

If he does not acquire the correct asset in the interim, then the analysis runs differently. If the beneficiary opted for the accounting remedy, the traditional view would be that the beneficiary would falsify the disbursement and the court would expect the trustee to provide substitutive compensation for all the monies that should have been in the fund. The trustee would either get credit for the asset or sell it and pay the money back. If the asset has fallen in value, the trustee would then have to reach into his own pocket to fully account.[82] This is the *Mothew*[83] approach to enforcing the stewardship, noted above.

But this begs a question: if the asset is materially identical in all respects, does enforcing the primary remedy mean he must sell it and acquire an identical asset? If a beneficiary opts to enforce the duty by bringing an action for equitable compensation and gets credit for the asset, then the asset need not be sold – monetary substitute will suffice.[84] This is the nature of equitable compensation, being a substitute for performance after a breach, rather than enforcing the performance in itself. Specific performance, rather than accounting, is the remedy that should be sought by a beneficiary seeking the asset *in specie*, and that is a matter for a different paper.

As a side note, in the opposite case where an acquired asset has become more valuable than the required asset, the beneficiary is unlikely to bring an action for account as he gets to keep the asset. If the trustee has sold it anyway, the beneficiary could trace the uplift into the proceeds – this is the orthodox position in the law of tracing and claiming.[85]

C. The final outcome

Irrespective of how one treats wrongly acquired assets, it is clear that on the basis of the submissions made to the Supreme Court, Redler would be liable for £2.5 million. This is despite the fact that half of that sum would have been lost even if Redler had not committed a breach of trust. In accordance with the reasons laid out above, the principles of equity demand that Redler account for the entirety of their breach of trust. This is true regardless of whether Redler needed to reconstitute the fund or pay the sum directly, a matter which Lord Toulson regarded as 'mere procedure'.[86] This is the approach which the Supreme Court ought to have taken if it wished to act consistently with the established accounting remedy argued for above.

Despite this being the correct outcome of the case, it does change the fact that the Supreme Court is unlikely to revert to the traditional remedy any time in the near future. The post-*AIB* landscape is the one equity practitioners will have to become accustomed to, and the exact limits of the decision must be ascertained quickly if parties are to enter into trust relationships with any degree of certainty.

---

[81] *Target Holdings Ltd v Redferns* [1994] 1 WLR. 1089 (CA), 1105-1106 (Peter Gibson LJ).
[82] Glister and Lee (n 59) 24-014.
[83] *Mothew* (n 39).
[84] Glister and Lee (n 59) 24-007.
[85] *Wright v Morgan* [1926] AC 788 (PC), 798.
[86] *AIB* (n 1) [91] (Lord Toulson).

7. *The law post-*AIB

A. Scope of the decision

To prevent a hard case from making bad law, a number of academics[87] have suggested that *AIB* is limited to its facts, displacing accounting and substitutive compensation in cases where the trust arises out of a contract for a contractual purpose, and that the traditional remedy still exists for trusts outside this scope. However, it is very difficult to square this narrow approach with their Lordships' speeches.

Lord Toulson notes, with a large degree of generality, that 'it would not in my opinion be right to impose or maintain a rule that gives redress to a beneficiary for loss which would have been suffered if the trustee had properly performed its duties.'[88] His subsequent statements do not in any way qualify or contradict this position. Quite the contrary, he then argues that accounting cannot exist alongside equitable compensation at all, and that compensation without reference to loss caused or gain made is 'penal'.[89] If one cannot have both in cases concerning traditional trusts, then there must either be equitable compensation requiring loss to be caused, or a remedy that is 'penal'. It is difficult to see how Lord Toulson intended to leave such a dichotomy open for other types of trust given his views on the traditional remedy. Lord Reed is less unequivocal, but while his judgment does draw some distinctions between the facts of *AIB* and the traditional remedy cases, he still states 'the aim of equitable compensation is to compensate: that is to say, to provide a monetary equivalent of *what has been lost as a result of a breach of duty* (emphasis added)'.[90]

More fundamentally, the suggested dichotomy is problematic, as was noted[91] after it was first mooted in *Target*.[92] If the distinction is based upon Lord Toulson's brief comment that that trusts '[arising] out of a contract'[93] demand different remedies than those arising gratuitously, then its scope is highly unclear. One may find trusts in traditional family settings that arise from agreement supported by consideration, and equally numerous commercial transactions are deliberately made by deed without consideration. If one includes deeds, then what of trust deeds used in the family trust context?

If the distinction is rooted in the commercial nature of the transaction rather than by its means of creation, then the boundaries of a commercial trust must be at least capable of being clear. But it is submitted this boundary is in no way clear: dividing the private and commercial is not conceptually straightforward and apt to cause confusion. The courts have arguably only done so with a simpler dichotomy in contract - between 'dealing as a consumer' and 'business'[94] - because it was necessitated by legislation.[95] Defining the boundary of 'commerce' is not a task the court should attempt to impose on itself without very good reason, and arguably their Lordships had avoided doing so.

Lord Toulson did opine that '[t]he contract defines the parameters of the trust', but this is not sufficient to prove that such contracts have differing rules of remedy without a substantial change to our understanding of contemporaneous liability. The underlying

---

[87] Ho (n 34); James Penner, 'Distinguishing Fiduciary, Trust, and Accounting Relationships' (2014) 8 J Eq 202, 225; McFarlane and Mitchell (n 21) 10-097.
[88] *AIB* (n 1) [62] (Lord Toulson).
[89] ibid [64] (Lord Toulson) discussed above in Part 5 Section D.
[90] ibid [138] (Lord Reed).
[91] Millett (n 6) 224-5; Davies (n 7) 687.
[92] *Target* (n 3*)* 435-6.
[93] *AIB* (n 1) [70] (Lord Toulson).
[94] Edwin Peel, *Treitel on The Law of Contract* (14th Ed, Sweet & Maxwell 2015) 7-052 to 7-054.
[95] Unfair Contract Terms Act 1977, s 1(3).

contract itself may of course modify the duties of the parties or limit or exclude liability for negligent breach.[96] Equally, policy may dictate that no other obligation ever arises in certain cases where contracts exist,[97] though one may question why the fact the trustee is being paid should not suggest he should be prepared to accept more extensive remedies. Regardless, if the contract does not modify liability expressly or negative any obligation, the fact the trust arises from contract should not itself change the remedy for breach of that obligation.

Cases which have addressed issues of concurrent liability often take for granted that the alternate obligation provides an advantageous remedy.[98] This is true even where the contract creates the obligation and does not exist independently. The fact that one contractually invites persons onto one's defective premises does not in itself entitle them to a different remedy in tort if they are injured by the defect,[99] even if they came onto the property for commercial purposes. Nor is a commercial bailee held to a different remedy because his possession arose from a contract, even if he faces a different standard of liability.[100] These are of course examples of legal relationships and not equitable ones, but Lord Toulson surely did not intend to lessen beneficiaries' protections alone and create a *sui generis* exception for trusts.

Furthermore, if the distinction were to be used, why should a 'commercial' trustee, who is being compensated for his services and usually will be insured against losses, be subject to a less extensive remedy than a 'traditional' trustee who is often unpaid and uninsured, merely because of the near inevitability that his obligation arises from a commercial contract, even when the contract itself is silent to exclusion or limitation of liability? The beneficiary is paying for the trustee to perform his obligations and so arguably has an equal if not greater claim to strong remedial protection. The contrary would be far more unjust and illogical than the traditional scheme applying to all trustees.

Given these factors, it is hard to see how the judgment does not apply to wrongful disbursements from all trusts, not just those similar to the one in *AIB,* and it is thus impossible to agree with the analysis confining *AIB* to its facts.

B. Clarity of the decision

AIB approves *Target* and its basis of recovery in no uncertain terms. Notwithstanding the argument about the scope of its application, there is little doubt it now provides the basis of recovery in cases of wrongful disbursement.

There is still, however, significant difficulty for future courts in determining how exactly the rules of equitable compensation now operate in regard to causation. The court did not in this case provide a great deal of guidance on the exact limits of causal liability in terms of factual and legal causation, beyond noting that the simple 'but for' test is necessary but will require further rules to supplement it.[101] *Target* itself made reference to a 'common sense'

---

[96] *Armitage v Nurse* [1997] EWCA Civ 1279, [1998] Ch 241. Solicitors cannot exclude such liability (SRA Handbook Chapter 1 Outcome 1.8), but this is the price that must be paid for being in a well-compensated but highly regulated profession, and given that solicitors must have malpractice insurance, it is certainly not grounds to weaken the default legal position.

[97] See, for example, *Marc Rich & Co AG v Bishop Rock Marine Co Ltd* [1995] UKHL 4, [1996] AC 211, 240 (Lord Steyn).

[98] *Henderson v Merrett Syndicates* [1994] UKHL 5, [1995] 2 AC 145, 194 (Lord Goff).

[99] Occupiers' Liability Act 1957, s 1. The common law position merely demanded a different (and incidentally) higher standard of *care* for contractual licensees than for 'invitees' – *Indermaur v James* (1866) LR 1 CP 274.

[100] *Coggs v Bernard* (1703) 2 Ld Raym 909.

[101] *AIB* (n 1) [64], [70], [73] (Lord Toulson), [117], [135]-[136] (Lord Reed).

approach,[102] but this could hardly be less helpful to trustees, beneficiaries or the lower courts alike going forward.

The failure of the court to provide clear rules with regards to causation and remoteness have dismayed even those who otherwise support this decision and Lord Browne-Wilkinson's judgment,[103] and it will suffice to say here that the uncertainty that accompanies it will no doubt generate many headaches for the courts going forward, especially as the Supreme Court applied only basic contractual principles, refusing to adopt the fully developed contractual or tortious rules of damages.[104] Future courts will have to determine what exactly 'causation' entails,[105] whether rules of remoteness are to apply, and if so, how.[106]

C. Commercial implications

The drawback, as with the restriction of any duty, is the diminution of the rights of the party to whom the duty is owed. The extent of accounting liability had acted as a prophylactic against such breaches. Now, however, provided the trustee gets lucky and outside circumstances render the loss following his wrongful disbursement inevitable, the trustee may escape liability for the breach of trust. As a result, beneficiaries can no longer be as certain that their property will not be negligently misapplied, which, as seen in *Target*, can involve behaviour that encompasses quite flagrant and egregious breaches. Contractual remedies may remain, depending on the terms of the settlement, but they are constrained by rules of causation and remoteness[107] and will not always be available. The exact commercial implications have been written about extensively[108] and are outside the scope of this article, whose focus is on the doctrinal implications for the law of trusts.

8. *Conclusion*

Overall, then, the decision in *AIB v Mark Redler* is not one that can be easily justified by normative reasons or on grounds of precedent. Further to this, its attack on the nature of the law of trusts is a worrying and misguided step. The *obiter* commentary of their Lordships on the boundary between equity and the common law should also be subject to strict scrutiny, and the courts should consider very carefully before adopting their reasoning.

This author concedes, nevertheless, that this case represents the judgment of the highest court in the land, and a departure from the decision in the immediate future is very unlikely. For now, the remedy for breach of trust by wrongful disbursement is left with a number of contractual principles shoehorned into it, despite their clash with fundamental principles of equity. Contrary to the analysis of numerous academics, most notably McFarlane and Mitchell,[109] it is unlikely that the decision will be confined to cases of commercial trusts given the generality expressed by their Lordships. However, due to the worrying implications of this decision, it is submitted that the courts should confine the decision to its facts as far as it is possible to do so, and that the issue should be revisited by the Supreme Court in the future.

---

[102] *Target* (n 3) 439 (Lord Browne-Wilkinson).
[103] Andreas Televantos and Lorenzo Maniscalco, 'Stay on Target: Compensation and Causation in Breach of Trust Claims' [2015] Conv 348, 349-351.
[104] *AIB* (n 1) [115]-[116], [136]-[137] (Lord Reed).
[105] ibid [136] (Lord Reed).
[106] ibid.
[107] See the application of the 'but for' test in contract in *South Australia Asset Management Corporation Respondents v. York Montague Ltd* [1996] UKHL 10, [1997] AC 191, 214 (Lord Hoffman).
[108] See, for example, Davies (n 7) 693-694.
[109] McFarlane and Mitchell (n 21) 10-097.

Rebates post-*Post Danmark II*

Sym Hunt*

Introduction

In *Post Danmark A/S v Konkurrencerådet* ('***Post Danmark II***'),[1] the European Court of Justice ('**CJEU**') gave a preliminary ruling which, *inter alia*, strongly indicated that, for the purposes of Article 102 of the Treaty on the Functioning of the European Union ('**TFEU**'),[2] retroactive rebates are treated with the same suspicion as exclusivity rebates, i.e. they are prohibited, unless objective justifications for the rebates are proven by the dominant undertakings.

The CJEU brought together the existing case law and adopted a two-stage, effects-based analysis, which, in theory, accords with the Commission's increasing emphasis on effects. However, this article will show that the conclusion of illegality arising from this allegedly effects-based analysis, in the context of retroactive rebates, is almost inevitable from the outset: the analysis considers effects in only a superficial manner. In that respect, the legal position of retroactive rebates is much more similar to that of exclusivity rebates than one might assume from: the General Court's *Intel Corp v Commission* ('***Intel***') judgment;[3] and, the general focus in EU jurisprudence on quantity-based rebates and exclusivity rebates having essentially pre-determined outcomes,[4] with other rebates being treated with greater nuance. This assimilation, in treatment, of retroactive and exclusivity rebates is logical, given the anticompetitive characteristics that this article will demonstrate are shared by both kinds of rebates. However, in order for the approach to be *desirable* and fair on dominant undertakings, the prospect of objective justifications needs to be taken seriously by the Commission and the CJEU.

This article will do five things:

i. outline the position on rebates before *Post Danmark II*;
ii. explain the *Post Danmark II* ruling in the broader context of the existing case law and approach to Article 102;
iii. illustrate how this judgment provides for a more consistent and logical approach to rebates to be more definitively and broadly adopted in the *Intel* appeal;[5]
iv. explain why the strictness of the approach which *Post Danmark II* signals is generally justified and yet in need of being offset by adequate consideration of objective justifications; and,
v. outline the questions that remain after the ruling.

* St John's College, Oxford. Comments are welcome (symeonhunt@outlook.com).

[1] Case C-23/14 *Post Danmark A/S v Konkurrencerådet* [2015] 5 CMLR 25.

[2] The facts of the case predated the entry into force of the Treaty of Lisbon, thus Article 82 EC was the applicable provision. However, the interpretation given in the preliminary ruling will undoubtedly apply to the successor provision, Article 102, and therefore that provision will be referred to throughout this note.

[3] Case T-286/09 *Intel Corp v Commission* [2014] 5 CMLR 9.

[4] See sections 1(A)-1(B) below.

[5] Case C-413/14 P *Intel Corporation v Commission* (21 June 2016).

1. *Pre*-Post Danmark II

A. Definitions and the tripartite categorisation

Before *Post Danmark II*, rebates had been, for legal purposes, divided into three categories:[6] quantity rebates; exclusivity rebates; and, rebates which do not fit neatly within either category (termed 'residual rebates', for present purposes). This tripartite categorisation was adopted by the General Court in *Intel*, and characterizes the approaches typically taken to different kinds of rebates.

Quantity rebates are where a customer enjoys a price reduction that is based solely on the volume of purchases.[7] Quantity rebates are 'generally considered not to have the foreclosure effect prohibited by [Art 102]',[8] because they are 'deemed to reflect gains in efficiency and economies of scale made by the undertaking in a dominant position'.[9]

Exclusivity rebates are where a customer enjoys a price reduction only if they obtain most or all of their requirements from the undertaking. Exclusivity rebates by a dominant undertaking are 'incompatible with the objective of undistorted competition within the common market'[10] because they are 'designed to remove or restrict the purchaser's freedom to choose his sources of supply and to deny other producers access to the market'.[11] The mischief behind the strict approach taken towards exclusivity rebates therefore appears to be protecting the commercial freedom of buyers and preventing foreclosure of (potential or actual) competitors.

The residual rebates are those by virtue of which a customer enjoys a discount without a direct condition of exclusive or quasi-exclusive custom, but 'where the mechanism for granting the rebate may also have a fidelity-building effect'.[12] These rebates are assessed in light of the criteria and rules governing the grant of the rebate,[13] and therefore, in theory, do not have the almost pre-determined outcome to which quantity rebates and exclusivity rebates are subject. The residual category includes target or 'conditional' rebates. Target rebates are where a customer enjoys a discount in light of purchasing a specified number of units. Within target rebates, Jones and Sufrin allude to four further descriptors,[14] which can be framed with regard to two further distinctions. First of all, target rebates may be *standardized*, or *individualized*. Standardized rebates are where all customers have the same target and discount; individualized rebates are where customers have different targets and/or discounts. Additionally, one can distinguish between *incremental* rebates, and *retroactive* rebates. Incremental rebates are where the discount applies only to purchase above the target. In contrast, retroactive rebates are where a customer is promised a discount *on all of their purchases* provided that they purchase a certain number of units from the undertaking. This is *retroactive* in the sense that the discount applies in respect of all the purchases which predate the meeting of the target. As such, they do not require exclusivity to a certain vendor, but are liable to *encourage* exclusivity, to an increasing extent as the customer nears the target.

Retroactive rebates are seen as concerning from a competition point of view, because they 'ensure that, from the point of view of the customer, the effective price for the last units is very low because of the *suction effect*',[15] which therefore means that competitors would have to offer extremely

[6] *Intel* (n 3) [74].
[7] Case 322/81 *NV Nederlandsche Banden Industrie Michelin v Commission* [1983] ECR 3461 ('***Michelin I***') [72].
[8] *Intel* (n 3) [75].
[9] ibid.
[10] ibid [77].
[11] ibid.
[12] ibid [78].
[13] ibid.
[14] Alison Jones and Brenda Sufrin, *EU Competition Law: Text, Cases, and Materials* (6th edn, OUP 2016) 435.
[15] *Case C-549/10 P Tomra Systems ASA v Commission* [2012] 4 CMLR 27 ('***Tomra***') [78].

low prices in order to gain the custom of the customer that is near to reaching its target. This is illustrated by the example below.

Assume that: X is a customer; Y is a dominant undertaking that supplies X; and, Z is attempting to compete with Y. Y promises X a 20% discount on all of its purchases, provided that it purchases at least 100 units. If each unit costs £1, then 100 units would cost £100 without the discount, and £80 with the discount. However, because the discount applies only retroactively, once 100 units have been bought, X continues to pay £1 per unit until that point. Before X has bought any units, Z is capable of offering as low a price by charging X 80 pence per unit. Once X has bought 50 units from Y for £50, its next 50 units will essentially cost only £30, because when it reaches 100 units, the rebate will be applied. As such, the average cost of the latter 50 units is 60 pence. Therefore, in order to compete on price, Z has to lower its asking price to 60 pence. In contrast, the average price received by Y would never be below 80 pence. Once X has bought 80 units from Y for £80, the next 20 units essentially cost X nothing. Therefore, in order to offer as favourable a deal, Z would have to give X 20 units for free; whereas, even when Y fulfils its promise of the 20% discount upon the target being reached, it has still received 80 pence per unit. When X has bought 99 units for £99, Z would have to pay X £19 *and* give X the unit for free, in order to match Y's offering. This is what is known as the *suction effect*: the closer X gets to the target, the more inclined it is to purchase from the rebate-offeror, and the less profitable (or greater loss-incurring) price competitors have to offer in order to compete. When this practice is implemented by dominant undertakings, and the figures are in millions, retroactive rebates can have an eliminatory effect on competition. As will become clear over the course of this article (and even simply by surveying the facts of *Post Danmark II*), retroactive rebates harbour some of the same anticompetitive characteristics as exclusivity rebates, and *can* be just as damaging to competition.

B. The legal analysis of rebates

As alluded to above, the outcome in respect of either a quantity rebate, or an exclusivity rebate, is clear from the outset, but the same could not, prior to *Post Danmark II*, be said for any residual rebates. It is crucial to note, however, that the positions in respect of quantity rebates and exclusivity rebates are not without qualification. For instance, in *Intel*, quantity rebates were said to be 'generally considered'[16] not to be anticompetitive; and, exclusivity rebates were seen as anticompetitive 'save in exceptional circumstances'[17] – the possibility of 'objective justification'[18] was recognized. As such, these outcomes are not absolute. The position of the residual rebates was considered by the CJEU in *British Airways Plc v Commission* ('***British Airways***'),[19] before the formal categorisation was made in *Intel*. It adopted a two-stage analysis ('**the two-stage test**', hereafter), in respect of (what are now) residual rebates.

The first stage is the determination of whether the rebates can produce an *exclusionary effect*, that is whether they are *capable* of both 'making market entry very difficult or impossible for competitors of the undertaking in a dominant position'[20] and 'making it more difficult or impossible for its co-contractors to choose between various sources of supply or commercial partners'.[21] In other words, the two elements of exclusionary effect concern difficulty of entry and customer freedom. This assessment will take account of 'all the circumstances',[22] particularly the criteria and rules governing the grant of the rebate, and 'the particular conditions of competition prevailing on the relevant market'. If the first stage is answered in the affirmative (i.e. there *is* deemed to be the capability of

[16] *Intel* (n 3) [75].
[17] ibid [77].
[18] ibid [81].
[19] Case C-95/04 P *British Airways Plc v Commission* [2007] ECR I-2331.
[20] ibid [68].
[21] ibid.
[22] *Post Denmark II* (n 1) [29].

exclusionary effect), the second stage entails consideration of whether there is an 'objective economic justification'[23] for the rebates.

C. Objective justifications

Objective justifications have their roots in the two-stage test explained above. In *British Airways*, the CJEU considered whether the exclusionary effect 'may be counterbalanced, or outweighed, by advantages in terms of efficiency which also benefit the consumer'.[24] On the facts, it was deemed not to be justified in that way, as is considered in section 2(C) below. However, the *British Airways* ruling was important for making clear the possibility of a defence to exclusionary effect, under Article 102.

The possibility of objective justifications was reiterated and developed in *Post Danmark v Konkurrencerådet* ('***Post Danmark I***'),[25] where the CJEU alluded to two kinds of objective economic justification: where the conduct is 'objectively necessary'; or, where the conduct may be 'counterbalanced, outweighed even, by advantages in terms of efficiency that also benefit consumers'.[26]

The notion of *objective necessity* was not made clear in *Post Danmark I*, but the Guidance Paper regarding the *enforcement priorities in applying [Art 102] of the EC Treaty to abusive exclusionary conduct by dominant undertakings* (the '**Guidance Paper**')[27] refers to conduct that is 'objectively necessary' to 'factors external to the undertaking'.[28] This is not particularly clear, and the only elaboration by the Commission is the example of 'health and safety reasons related to the nature of the product in question', and reiteration that dominant undertakings must not take steps on their own initiative to exclude products which it regards as dangerous or inferior. The scope of this version of the defence is, therefore, unclear – although, a possible use of it will be considered in section 4(C)(I) below.

The *efficiencies* version of the defence was more comprehensively explained by the CJEU, in *Post Danmark I*. It was said that 'it is for the dominant undertaking to show that the efficiency gains likely to result from the conduct under consideration counteract any likely negative effects on competition and consumer welfare in the affected markets, that those gains have been, or are likely to be, brought about as a result of that conduct, that such conduct is necessary for the achievement of those gains in efficiency and that it does not eliminate effective competition, by removing all or most existing sources of actual or potential competition'.[29] The Guidance Paper endorses this defence in a similar manner,[30] but refers to there being 'no net harm to consumers [that] is likely to arise',[31] and requires that the efficiency gains 'outweigh'[32] (rather than counteract) any likely negative effects on competition and consumer welfare. This has been said by Nazzini[33] to be inconsistent with *British Airways*, and to be too high a threshold. This may well be true, but it is certainly possible and realistic that some conduct may provide the requisite neutral or positive effect on consumer welfare; indeed, that is something that it is reasonable to encourage dominant undertakings to do.

---

[23] *British Airways* (n 19) [69].
[24] ibid [86].
[25] Case C-209/10 *Post Danmark v Konkurrencerådet* [2012] ECR I-0000.
[26] ibid [41].
[27] Commission, 'Guidance on the Commission's enforcement priorities in applying Article 82 of the EC Treaty to abusive exclusionary conduct by dominant undertakings' (Communication) (2009/C 45/02).
[28] ibid [29].
[29] *Post Denmark I* (n 25) [42].
[30] Guidance Paper (n 27) [30].
[31] ibid.
[32] ibid.
[33] Renato Nazzini, *The Foundations of European Union Competition Law: The Objective and Principles of Article 102* (OUP 2011) 308.

Whilst the objective justifications defence appears to be narrow, it is available, and the full extent of its reach remains to be seen. It is important to note the preface, in the *Post Danmark I* description of the versions of the defence, of '[i]n particular', i.e. it was recognized that other versions may exist (this will be referred to as 'the "in particular" caveat'). As will be demonstrated in section 4(C)(I) below, there is already a third possible version of the defence waiting to be realized, and it should not be assumed that the current versions of the defence amount to an exhaustive list of the possible circumstances in which the defence may be attained. This article will use the term 'objective justifications' to refer to this defence in its entirety (thereby including the two recognized versions of the defence, and other possible versions of the defence).

2. *Post Danmark II*

A. Factual background and the preliminary ruling

The Konkurrencerådet (the Danish Competition Council) found that Post Danmark A/S ('**Post Danmark**') was an 'unavoidable trading partner'[34] on the relevant market for the 'distribution of bulk mail',[35] with a market share of over 95%. During the relevant period, the market for the distribution of bulk mail was subject to high access barriers due to economies of scale, and a statutory monopoly that accounted for 70% of total demand. Therefore, Post Danmark clearly held a dominant position.[36] The only 'serious competitor' on the market was Bring Citymail A/S ('**Bring Citymail**'),[37] which, when active, delivered direct advertising mail in a service available to approximately 40% of the relevant households.[38]

Post Danmark implemented a rebate scheme in respect of direct advertising mail. Though this is the segment of the market in which Bring Citymail sought to compete with Post Danmark, it should be noted that the implementation of a standardized rebate scheme was not a reaction to Bring Citymail's entry, as the rebates began in 2007, almost four years before Bring Citymail's entry into the market. The rebates applied to mailings which were: presented in batches of at least 3,000 at a time; and, aggregated at least 30,000 letters per year, or represented a minimum annual gross value which was approximately €40,200[39] – crucially, the rebate applied retroactively to those who met the target.

Having suffered losses in the region of €67m, Bring Citymail withdrew from the Danish market in 2010, but had lodged a complaint with the Konkurrencerådet in 2009. Their complaint precipitated the proceedings that eventually led to the preliminary ruling. The Konkurrencerådet found that Post Danmark had abused its dominant position, by applying rebates which 'had the *effect* of *tying customers* and *'foreclosing' the market*, without being able to substantiate the *efficiency gains* that might have benefited consumers and *neutralised those rebates' restrictive effects on competition*'.[40] The Konkurrenceankenævnet (the Competition Appeals Tribunal) upheld that decision, and Post Danmark consequently brought the case before the Sø- og Handelsretten (the Maritime and Commercial Court). Due to the uncertainty of the criteria which determine whether a rebate scheme is *capable of having an exclusionary effect*, the Sø- og Konkurrenceankenævnet referred a number of questions to the CJEU under Article 267 TFEU.

The questions referred sought a significant amount of clarification surrounding the application of Article 102 to rebates, thus it is no surprise that the CJEU gave such a comprehensive statement of the law. The CJEU reformulated the questions referred to, with the effect of presenting itself with,

[34] *Post Denmark II* (n 1) [14].
[35] ibid [13].
[36] ibid [14].
[37] ibid [11].
[38] ibid [10].
[39] ibid [7].
[40] *Post Denmark II* (n 1) [13] (emphasis added).

inter alia, an opportunity to 'clarify the criteria that are to be applied in order to determine whether a rebate scheme, such as that at issue in the main proceedings, is liable to have an exclusionary effect on the market contrary to [Article 102 TFEU]'.[41] This could be argued to have slightly widened the inquiry presented by the Sø- og Handelsretten, but this widening is welcome as it provides greater certainty in the general interpretation of an area which has seen a number of rulings[42] in recent years.

B. The CJEU's judgment

The CJEU characterized the rebates as 'standardised', 'conditional', and 'retroactive'.[43] It was noted that the rebates could not be regarded as 'simple quantity rebate[s]', nor as 'loyalty rebate[s]',[44] and therefore was a residual rebate. Interestingly, *Intel* was not referred to at all by the CJEU. Given the huge interest which has surrounded the *Intel*, one might have expected some reference to it by the CJEU, even if just an endorsement of the somewhat formulaic tripartite distinction used by the General Court in *Intel.*[45] As shall become clear in this article, the assimilation of retroactive and exclusivity rebates undermines the tripartite division in its current form. Nonetheless, it should be noted that the two-stage analysis was applied was used *because* of the circumstances, i.e. because the rebates were not clearly either volume- or exclusivity- based. As such, one should not hastily assume that the tripartite division has been abandoned. Indeed, Sidiropoulos has claimed that the CJEU confirmed the General Court's approach[46] – in lieu of explicit use of the tripartite distinction, this seems premature. As such, it will be interesting to see if the tripartite distinction will be reaffirmed in the *Intel* appeal,[47] which was heard in June 2016 – this will be considered in section 5 below.

Applying the two-stage test, the CJEU found that the rebate scheme produced an anti-competitive exclusionary effect.[48] Aggravating factors in reaching that conclusion included the fact that the rebates applied retrospectively, the length of the reference period (one year), and the extent of Post Danmark's dominance (bearing in mind its statutory monopoly and significant market share). The second stage of the test was reiterated, but the CJEU did not pass judgment on the satisfaction of the objective justification defence.

C. Chan's 'virtually unavailable justifications'

Sunny Chan has argued that, in light of the CJEU's analysis in *Post Danmark II*,[49] retroactive rebates are *per se* unlawful, rather than retroactivity simply being a contributory factor. Furthermore, Chan claims that *objective economic justifications* are '*virtually* unavailable',[50] in the case of retroactive rebates, in light of failed attempts to justify rebates making reference to diminishing marginal utility and economies of scale. It is submitted that, whilst Chan is right to suggest that it is difficult for undertakings to avail themselves of the objective justifications, his claim of economies of scale being a virtually unavailable justification is exaggerated and presumptuous. As such, it is too soon to say that retroactive rebates are *per se* unlawful – they are subject to the same *qualified per se* prohibition that

---

[41] *Post Denmark II* (n 1) [21].
[42] *British Airways* (n 19); Case T-66/01 *Imperial Chemical Industries Ltd v Commission* [2010] ECR II-2631; *Intel* (n 3); *Tomra* (n 15).
[43] *Post Denmark II* (n 1) [23]-[25].
[44] ibid [28].
[45] *Intel* (n 3) [75]-[78].
[46] Konstantinos Sidiropoulos, 'Post Danmark II: A Clarification of the law on Rebates under Article 102 TFEU' (*European Law Blog*, 11 December 2015) <http://europeanlawblog.eu/?p=3013> accessed 21 September 2016.
[47] n 5.
[48] *Post Denmark II* (n 1) [42].
[49] Sunny SH Chan, 'Post Danmark II: per se unlawfulness of retroactive rebates granted by dominant undertakings' (2016) 37(2) ECLR 43.
[50] ibid [49] (emphasis added).

exclusivity rebates are. The extent of this qualification is unclear, but it will be submitted that the qualification must be taken seriously in order for the rule to operate fairly to dominant undertakings.

Chan cites *Michelin I*,[51] where the CJEU clearly refuted wishes to 'sell more' or 'spread production more evenly'[52] as economic justifications. It is clear from the judgment, and its explanation by Chan,[53] that the CJEU has rejected this economic justification, in the context of retroactive rebates. However, Chan misrepresents the CJEU's rejection of this justification – he claims that diminishing marginal utility has been rejected in *Michelin I*. But this is not the case: seeking to spread production more evenly is completely different from taking note of the correlation between units purchased, and elasticity of demand. As such, whilst *Michelin I* does show a rejection of the objective justifications of wishes to sell more or spread production more evenly, undertakings *can* allude to diminishing marginal utility as a justification.

In respect of economies of scale, Chan draws upon *British Airways*.[54] Chan is correct to state that economies of scale did not prevail as an objective justification, however, as he recognizes, this was because the CJEU was unwilling to substitute its own assessment of market data for that of the Court of First Instance[55] – two conclusions follow from this: that the Court of First Instance deemed the justification not to have been demonstrated on the facts; and, that the CJEU was unwilling to reconsider the assessment of data that resulted in the former conclusion by the Court of First Instance. The importance of these two conclusions is that they show that Chan misleads us by suggesting that the CJEU rejected the economies of scale justification in principle. This is not the case – the Court of First Instance did not deem the justification to have been demonstrated *on the facts*, and the CJEU was unwilling to substitute its assessment of those facts, meaning that any conclusion of an error of law would not have been likely. The possibility of economies of scale as a justification was not at all ruled out by *British Airways*. On the contrary, the Court of First Instance implicitly recognized its existence as a justification, in that it was willing to consider whether the facts demonstrated that the rebate scheme 'allow[ed] [BA] to reduce its costs'.[56] This suggests that *if BA had shown economies of scale by virtue of reduced costs flowing from the rebate scheme*, then there would have been the possibility of an objective justification. Furthermore, the Commission has shown clear attempts[57] to undertake a more economic analysis in the context of rebates in infringement Decisions since *British Airways*, thus it is submitted that such a justification is far from *virtually unavailable*.

Furthermore, Chan claims that '[t]he virtual impossibility of justifying a retroactive rebate is borne out by the fact that all dominant undertakings in *British Airways*, *Tomra*, and *Post Danmark II* did not succeed in proffering a justification for their retroactive rebates'.[58] However, the mere fact that these undertakings failed to provide such a justification does not in itself demonstrate its 'virtual impossibility' – it could just mean that no such justification existed on the facts. Indeed, it seems unlikely that an undertaking with a 95% market share and a statutory monopoly over most of that market could justify a retroactive rebate scheme with a long reference period. Similarly, Tomra exercised a complex structure of exclusivity agreements, quantity commitments and fidelity-inducing discounts – this demonstrated a patchwork of agreements geared towards foreclosure. British Airways based its rebates on increases in sales from one year to the next, with the possible result of such a

[51] *Michelin I* (n 7).
[52] *Post Denmark II* (n 1) [85].
[53] Chan (n 49) 49.
[54] ibid.
[55] *British Airways* (n 19) [88].
[56] ibid [291].
[57] See, in particular: *Prokent-Tomra* (Case COMP/E-1/38.113) Commission Decision 2008/C219/12 [332]; and, the enormous *legal and economic assessment* that the Commission undertook in *Intel* (Case COMP/37.990) Commission Decision 2009/C227/07.
[58] Chan (n 49) 49.

scheme being an exponential year-on-year suction effect – a rebate scheme of this kind does *not* seem correlative to or necessary as to meet high fixed costs. As such, it is respectfully submitted that the possibility of retroactive rebates being justified under the second limb of the *British Airways* test is not precluded from the CJEU's approach, nor does the existing case law demonstrate that such a justification is impossible, or even virtually available. The prospect of attainable objective justifications is further considered in section 4(C) below.

3. *The assimilation of retroactive rebates and exclusivity rebates*

The two-stage analysis adopted in *Post Danmark II* is, as has been noted, nothing new. Stemming from *British Airways*, it was also alluded to by the General Court in *Intel*, in the context of exclusivity rebates.[59] This led to Nicholas Petit's claim that, following *Intel*, exclusivity rebates are subject to a 'modified *per se* prohibition rule'.[60] The *modification,* which Petit referred to, is the qualification that objective justifications can prevent an exclusivity rebate from infringing Article 102 (i.e. the second limb of the two-stage test).

This qualified prohibition approach is not exclusive to rebates, Article 102, or even competition law as a whole. We see it in respect of each of the four freedoms: in respect of goods, Article 34 is qualified by Article 36; in respect of services and establishment, Article 49 is qualified by Article 52 and Article 56 is qualified by Article 62; in respect of capital, Article 63 is qualified by Article 65; and, in respect of persons, Article 27 of Directive 2004/38 qualifies the various rights of movement. Within competition law, Article 102's sister prohibition, Article 101(1), is qualified by Article 101(3). The approaches in respect of the four freedoms and Article 101 are analytically and formally different to that of Article 102, in that the equivalent justificatory defences, in respect of the former, render conduct outside the scope of the prohibition (or, in the case of Article 101, render the prohibition *inapplicable*).[61] However, the premises are the same: behaviour that is factually or assumedly unlawful can escape the finding of an infringement of EU law in all of these contexts, by virtue of meeting certain conditions; Article 101 and 102 are both familiar with qualified rules.

Furthermore, in the contexts of both Article 101 and Article 102, certain kinds of behaviour have been seen as automatically abusive. Within Article 101, restrictions 'by object' have traditionally been assumed to automatically be abusive (e.g. price-fixing[62]), though it is worth noting both the potential significance of the *de minimis* Notice[63] in reducing the scope of the applicability of Article 101 to 'object' restrictions, as well as the recent dilution by the CJEU of the analysis taken in respect of restrictions by object.[64] Within Article 102, prices below average variable cost have been seen as necessarily abusive.[65] It is, therefore, nothing novel for certain conduct to be automatically regarded as abusive. Similarly, Article 101(3) will seldom apply to object or hard-core restrictions, and *objective justifications* cannot realistically be proven in cases of selling below average variable cost,[66] so it would

---

[59] *Intel* (n 3) [80]-[81].
[60] Nicolas Petit, 'Intel, Leveraging Rebates and the Goals of Article 102 TFEU' (2015) 11(1) Euro C J 26.
[61] Consolidated versions of the Treaty on European Union and the Treaty on the Functioning of the European Union [2016] OJ C202, Article 101(3).
[62] Case 8/72 *Vereeniging Van Cementhandelaren v Commission* [1972] ECR 977; the Notice on agreements of minor importance which do not appreciably restrict competition under Article 101(1) of the Treaty on the Functioning of the European Union (C(2014) 4136) [2.1.1].
[63] Notice on agreements of minor importance which do not appreciably restrict competition under Article 101(1) of the Treaty on the Functioning of the European Union *(C(2014) 4136).*
[64] See, in particular: Case C-32/11 Allianz Hungária v Gazdasági Versenyhivatol [2013] 4 CMLR 25 [36]; Case C-67/13 Groupement des Cartes Bancaires v Commission [2014] 5 CMLR 22 [53]; Case C-172/14 ING Pensii [2015] 5 CMLR 15 [31]; Case C-286/13 Dole Food v Commission [2015] 4 CMLR 16 [117]; Case C-345/14 SIA 'Maxima Latvija' v Konkurences padome [2016] 4 CMLR 1 [17]; and, Case C-373/14 Toshiba Corp v Commission [2016] 4 CMLR 15 [26]-[29].
[65] Case 62/86 *AKZO Chemie BV v Commission* [1991] ECR I-3359 [71].
[66] *Post Denmark I* (n 25) [27].

be nothing revolutionary for retroactive rebates by dominant undertakings to be *per se* unlawful (and thereby excused from the finding of an infringement).

What the examples above seek to demonstrate is that a presumed prohibition rule is far from unheard of in EU competition law, and the same is true for a *qualified* prohibition rule. Focusing on rebates, and the effect of *Post Danmark II*, it becomes clear that this is precisely what the CJEU has, in effect, fostered in its judgment. It is crucial, at this point, to note that the *analysis* of exclusivity rebates and retroactive rebates *does* still superficially differ following *Post Danmark II*. The General Court, in *Intel*, was explicit in stating that 'an analysis of the circumstances of the case aimed at establishing a potential foreclosure effect' is not necessary in order to find that exclusivity rebates are abusive.[67] In contrast, the General Court noted, in the context of *third category* rebates, that the criteria and rules need considering,[68] and this was supported by the CJEU in *Post Danmark II*,[69] with 'all the circumstances' being relevant. As Chan correctly recognizes, 'all the circumstances' concerns whether or not the rebate is retroactive, with the result that retroactivity renders a rebate unlawful.[70] In this respect, it is open to question whether retroactive rebates really are treated differently to exclusivity rebates. The *process* by which the qualified *per se* prohibition comes about is claimed to be different, however, the *end result* is typically not. This is because the objective justification defence provides, in the context of exclusivity rebates, the same opportunity that (but for the narrow application which Chan highlights) would arise during the effects analysis of 'all the circumstances', in respect of retroactive rebates. A more logical approach would surely be to admit that retroactive rebates are seen in the same light as exclusivity rebates, and are treated as such. This would mean applying the same qualified *per se* prohibition rule.

The claim that retroactive rebates are now subject to, in effect, a qualified *per se* prohibition rule alongside exclusivity rebates is hardly a shocking or striking conclusion – both situations concern rebates which have traditionally been seen as abusive under Article 102. What the conclusion does demonstrate is the CJEU having a more uniform approach to rebates which are conducive to building fidelity in an anticompetitive manner. Retroactive (but non-exclusive) rebates and exclusivity rebates differ in that fidelity is not a *condition* (but, rather, a likely result) of the former, whereas it is a condition (and, indeed, the very premise) of the latter. However, the feared effect (and frequently common intention) of both is similar: exclusionary effects based on either the encouragement or requirement of buying (almost or entirely) exclusively from the dominant undertaking. In other words, the mischief behind any finding of retroactive or exclusive rebates as abusive under Article 102 is preventing arrangements whereby undertakings are put under overwhelming commercial pressure to purchase from the dominant undertaking *rather than its competitors*, *because of* the rebate scheme operated by the dominant undertaking. As such, it is logical to have a similar approach, with the same outcome, to both kinds of rebates. These rebates are distinct from the nature of quantity rebates, in that there is clear and independent commercial pressure in the context of exclusivity rebates (the condition of exclusivity) and retroactive rebates (the suction effect), whereas there is no such pressure in the context of quantity rebates.

Post Danmark II was important in nurturing the common approach: the CJEU's analysis emphasized, when considering 'all the circumstances', the retroactive nature of the rebates[71] in almost as damning a manner as the General Court did in *Intel*, in respect of exclusivity rebates.[72] This is why the different form of analysis does little to distinguish or affect the likely conclusions that will be drawn in cases of retroactive rebates, and exclusivity rebates, respectively. This generally strict view of

[67] *Intel* (n 3) [80].
[68] ibid [78].
[69] *Post Denmark II* (n 1) [29].
[70] Chan (n 49) [47]-[48].
[71] *Post Danmark II* (n 1) [32]-[33].
[72] *Intel* (n 3) [77].

the CJEU is consistent with the previous case law[73] and, as the following section will demonstrate, capable of being appropriate given the similarities shared by exclusivity rebates and retroactive rebates.

4. *Qualified* per se *prohibition of retroactive rebates – the right approach?*

This section will analytically consider the appropriate role of the effects of the allegedly anticompetitive conduct. As explained below, the Commission is increasingly considering the effects of the conduct in question, so as to determine whether it amounts to an exclusionary abuse; on the other hand, the Courts have indicated that the effect on competition is not as crucial an issue. If the Courts are not concerned with the effects on competition, for the purposes of reaching the conclusion that there is exclusionary abuse (under the first stage of the two-stage analysis), then it is particularly important that the lack of anticompetitive effects can still prevent the finding of an infringement of Article 102. In other words, this article argues that retroactive rebates are subjected to a qualified *per se* prohibition – if this is to be justifiable, then the qualification must be more than illusory, i.e. the objective justifications must not be *virtually unavailable*, and must be taken seriously by both the Commission and the European Courts.

A. The effects-based approach

In 2009, the Commission published the aforementioned Guidance Paper.[74] In respect of price-based exclusionary conduct (such as rebates), the Commission stated that it will normally intervene 'where, on the basis of cogent and convincing evidence, the allegedly abusive conduct is likely to lead to anti-competitive foreclosure'.[75] Various factors were outlined as to how such a determination will be made, and it was stated the assessment 'will usually be made by comparing the actual or likely future situation in the relevant market (with the dominant undertaking's conduct in place) with an appropriate counterfactual, such as the simple absence of the conduct in question or with another realistic alternative scenario, having regard to established business practices'.[76] This shows that the Commission is interested in the impact that the practice in question has had on the market, i.e. it has moved towards an *effects-based approach*.

It must be noted that the Commission states that '[i]f it appears that the conduct can only raise obstacles to competition and that it creates no efficiencies, its anti-competitive effect may be inferred'. There is a hugely important distinction between this, and the approach that is suggested by this article. This article suggests *prima facie* prohibition, with the (genuine) possibility of objective justifications; whereas, the inference suggested by the Commission in the quote above is one of *automatic* and *unqualified* prohibition. This article is concerned not with the Commission's claim of some practices possibly being capable of having their competitive intent inferred, but, rather, with the general effects-based approach which the Commission outlines in the Guidance Paper (as explained in the previous paragraph).

B. The CJEU's neglect of the effects-based approach for the purposes of establishing exclusionary effect

Whilst the Commission has indicated its intent to pursue a more effects-based approach, the Courts have, thus far, been cursory in any such analysis. Indeed, *Intel* and *Post Danmark II* are ideal examples of this.

In *Intel*, the General Court was dismissive of the need for an analysis of effects, particularly in the context of exclusivity rebates. The General Court stated that 'the question whether an exclusivity rebate can be categorized as abusive does not depend on an analysis of the circumstances of the case

[73] See the list in (n 42).
[74] n 27.
[75] Guidance Paper (n 27) [20].
[76] Guidance Paper (n 27) [21].

aimed at establishing a potential foreclosure effect';[77] and, that the as-efficient-competitor test (another effects-based element of the Guidance Paper) would not 'rule out the possibility that [market access] has been made more difficult'.[78] This shows the General Court being disinterested in effects-based analysis, in instances where the anticompetitive nature of conduct can be presumed.

In *Post Danmark II*, the CJEU confirmed that the as-efficient-competitor test 'must be regarded as one tool amongst others for the purposes of assessing whether there is an abuse of a dominant position in the context of a rebate scheme'.[79] The CJEU supported AG Kokott's claim that there was nothing in [Article 102 to 'support the inference of any legal obligation requiring that a finding to the effect that a rebate scheme operated by a dominant undertaking constitutes abuse must always be based on a price/cost analysis'[80] such as the as-efficient-competitor test.

It is clear from *Intel* and *Post Danmark II* that the effects-based approach is unlikely to significantly impact upon the jurisprudence of the Courts. For that very reason, it is crucial that dominant undertakings are given proper opportunity to justify behaviour, on the basis of it not being anticompetitive.

C. The answer to the neglect of the effects-based approach: taking objective justifications seriously

The neglect, by the Courts, of the effects-based approach can be (and has rightly been)[81] criticized, but the purpose, in this article, is to *explain* the current approach of the Courts, and propose how it can be further developed into a justifiable and clear approach to rebates. In other words, this article recognizes that the CJEU has made clear its disinterest with effects *for the purposes of demonstrating exclusionary effect*, and therefore considers the more open question of objective justifications, and how this stage in the two-stage analysis can be utilized to provide a more coherent framework for assessing rebates.

It is submitted that the place for consideration of effects is the determination of whether objective justifications exist (i.e. the second stage in the two-stage test), rather than whether a rebate prima facie has an exclusionary effect. The reason for this is that it is a more certain, predictable manner of analysis. Dominant undertakings know that retroactive or exclusivity rebates will be seen as unlawful *save for* objective justifications, rather than having to rely on an inconsistent (as between the Courts and the Commission),[82] and therefore uncertain, approach to the relevance of effects to the establishment of an exclusionary effect. As such, they can, as a matter of course, consider what objective justifications might be raised for such a practice, before embarking on the rebate scheme. A quite legitimate complaint against this would be that it appears to reverse the burden, given that '[i]t is incumbent upon the dominant undertaking to provide all the evidence necessary to demonstrate that the conduct concerned is objectively justified'.[83] However, it is submitted that the current situation has, in essence, already seen that burden reverse – formally recognising this and more seriously considering objective justifications can actually be *advantageous* to dominant undertakings. The benefit it has is that undertakings would not be under the illusion of their rebates being truly assessed in all of the relevant circumstances before a conclusion as to their exclusionary effect is determined – *Post Danmark II* and *Intel* demonstrate that exclusivity and retroactive rebates are both subjected to essentially pre-determined outcomes. What this article is proposing is that these pre-determined outcomes (in respect of the first stage of the analysis) be more openly acknowledged and

[77] *Intel* (n 3) [80].
[78] ibid [150].
[79] *Post Danmark II* (n 1) [61].
[80] ibid, Opinion of AG Kokott [61].
[81] Patrick Rey & James Venit, 'An Effects-Based Approach to Article 102: A Response to Wouter Wils' (2015) 38(1) World Competition 3, 10.
[82] ibid; section 4(B) above.
[83] Guidance Paper (n 27) [31].

allowed to form part of a genuine exercise of the purported two-stage analysis, whereby objective justifications are taken seriously.

It has been demonstrated in section 2(C) above that the prospect of objective justifications is by no means excluded. Section 2(C) put forward the claim that the fact that objective justifications were not, in the view of the Commission or the General Court/CJEU (as applicable), demonstrated in the likes of *British Airways*, *Tomra*, *Intel*, and *Post Danmark II* does *not* mean that they are virtually unavailable. It does not follow, however, that they are *sufficiently* available. Section 2(C) outlined the two currently established varieties of objective justifications: *objective necessities*; and, *efficiencies*. The former of these is unclear, and the latter is very narrow. In order to take objective justifications seriously, and thereby justify the qualified *per se* approaches taken to exclusivity and retroactive rebates, it is essential that objective justifications are construed in a more predictable and broad manner. This section will consider how objective justifications might be used in practice by dominant undertakings to justify (in particular) retroactive rebates, alluding to: meeting the competition; economies of scale; and, high fixed costs.

I. Meeting the competition

Meeting the competition has typically been seen in the context of selective price cuts. In *France Télécom*, it was held that 'the fact that an undertaking is in a dominant position cannot deprive it of the right to protect its own commercial interests if they are attacked and such an undertaking must be allowed the right to take such reasonable steps as it deems appropriate to protect those interests', with the caveat that 'such behaviour cannot be countenanced if its actual purpose is to strengthen this dominant position and abuse it'.[84] It could well be argued that where a dominant undertaking operates a retroactive rebate so as to protect its commercial interests, but does not have the *purpose* of strengthening or abusing their dominant position, then the defence should be available. As Jones and Sufrin have noted, this gains support from the reference, in the midst of the discussion of justifications in *Post Danmark I*, to *United Brands*,[85] where it was accepted that 'fact that an undertaking is in a dominant position cannot disentitle it from protecting its own commercial interests if they are attacked, and that such an undertaking must be conceded the right to take such reasonable steps as it deems appropriate to protect its said interests'.[86]

In the context of exclusivity rebate, it is understandable that the Commission would not allow the defence of meeting the competition: one cannot truly be said to *meet* the competition by eliminating it. However, in respect of retroactive rebates, it is much more plausible that one might meet the competition without eliminating it – there is a clear difference in terms of anticompetitive intent. Of course, if the effect of the rebate *was* to eliminate competition, then the defence would fail;[87] however, in other cases, the meeting of competition or protection of commercial interests could encapsulate a basis of objective justification. In the context of *Post Danmark II*, the condemned scheme predated Bring Citymail's entry into the market – if it had been a reaction to such a market entry, then one could envisage the *meeting the competition* defence being at least arguable.

The taxonomy of this as a separate justification or falling within *objective necessity* is less important than the *substance* of what it can exist in, but it is worth brief consideration. Jones and Sufrin take the view that the CJEU considers that protection of commercial interests may be 'another type of justification rather than a separate defence',[88] i.e. it is a third version of the objective justification defence. However, given the reference to health and safety needs, in the Guidance Paper,

[84] Case C-202/07 P *France Télécom v Commission* [2009] ECR I-02369 [46].
[85] Case 27/76 *United Brands Company and United Brands Continentaal BV v Commission* [1976] ECR 00425.
[86] ibid [189].
[87] See section 1(C) above.
[88] Jones and Sufrin (n 14) 371.

it might be most logical to recognize that the tenor of the *protection of commercial interests* approach is different to that of *objective necessity*, as understood by the Commission. Further, the 'in particular' caveat could be said to provide a basis for it being recognized as a separate, additional justification. However, if the Commission was to stick religiously to the defences as laid down in the Guidance Paper, then there is a risk that the meeting the competition defence would fail if treated as beyond the scope of the two defences referred to (though such an approach would be worthy of criticism, particularly given the 'in particular' caveat). It remains to be seen how the Commission would, post-Guidance Paper, deal with a defence based upon the protection of commercial interests, but given that the CJEU has endorsed it as a possible defence, the Commission would be foolish to dismiss, without due consideration, the defence, if raised.

Ethically speaking, one certainly could qualm the existence of a defence of *meeting the competition*. Indeed, Nazzini sees the concept as 'plainly absurd', in that it apparently legalizes predation.[89] This argument is not without merit, but the question of whether or not a defence is desirable is not within the ambit of this article – the important issue for present purposes is which defences are recognized within Article 102 and are, therefore, potentially available in the context of rebates. The discussion above demonstrates that meeting the competition is a recognized objective justification.

II. Economies of scale

Another possible basis for justification which the Commission highlights in the Guidance Paper is that where the rebate system is 'indispensable to the realisation of those efficiencies'.[90] In the context of retroactive rebates, this might refer to instances whereby the dominant undertaking needs to operate at a certain level of output in order to realize certain economies of scale, the benefit of which is then passed on to customers. In order to achieve this level of output, great initial investment may be needed, meaning that the discounts cannot be applied from the outset, but can be applied retrospectively, once the economies of scale have been realized. The Commission has previously been clear that '[s]uch cost savings need to be substantiated' and cannot merely be '[g]eneral remarks about transaction cost savings or general claims of better production planning'[91] – this justification is therefore not satisfied by spurious *ex post* claims, *but* it certainly can be satisfied by the dominant undertaking who, aware of the *prima facie* prohibition on retroactive rebates, only undertakes such a scheme having analysed and determined the existence and extent of these cost savings. This shows precisely how the qualified *per se* prohibition can provide greater certainty to dominant undertakings. Indeed, an undertaking which could adduce evidence in form of analysis pre-dating the introduction of the rebate would be in a strong position to demonstrate both the existence of indispensable savings, and the requisite causal link.[92]

The Commission considers, in the context of citing 'transaction-related cost advantages' as efficiencies, that 'incremental rebate schemes are in general more likely to give resellers an incentive to produce and resell a higher volume than retroactive rebate schemes'.[93] This (unsurprisingly) *implies* a preference for incremental schemes over retroactive schemes, but in no way suggests that retroactive schemes would be incapable of demonstrating such an efficiency.

III. High Fixed Costs

[89] Nazzini (n 33) 302.
[90] Guidance Paper (n 27) [30].
[91] DG Competition discussion paper on the application of Article 82 of the Treaty to exclusionary abuses (December 2005) [173]. Although this predates the current Guidance Paper, it is useful for showing the attitude taken by the Commission when consulting in respect proposals for the current Guidance Paper.
[92] Listed as a precondition in Guidance Paper (n 27) [30].
[93] ibid [46].

A similar idea to the economies of scale argument is that noted by Colomo,[94] whereby a retroactive quantity rebate scheme over a short period may be credible in an industry with high fixed costs. In such a circumstance, it would be reasonable to argue that such rebates should not be seen as an infringement under Article 102, and it would be realistic to demonstrate an objective economic justification for the rebates: namely, that the undertaking is keen to provide its customers with reasonable prices, but is unable to do so until the end of the short reference period, because of high fixed costs.

It should be pointed out that, whilst one might reasonably assume that rebates offered by an airline could conceivably be justified on this basis, on the facts of *British Airways* the exponential aspect[95] of the condition of the rebates suggests that there was no bona fide intention to deal with high fixed costs. Rather, it was an attempt to guarantee increased sales from each customer each year with no necessary correlation to the year-on-year fixed costs: the rebate was 'calculated on a sliding scale, based on the extent to which a travel agent increased the value of its sales of BA tickets, and subject to the agent's increasing its sales of such tickets from one year to the next'.[96] If a rebate was, instead, based upon a set target volume of sales, and was of a percentage or amount which was unaffected by the extent to which the target is exceeded, then the argument that retroactive rebates was necessary and justifiable, in light of high fixed costs, becomes much more plausible.

D. Interim conclusions

Post Danmark II shifts the treatment of retroactive rebates into the qualified *per se* prohibition rule which has been seen in the context of exclusivity rebates, in *Intel*. This is reflective of the fact that neither kind of rebate is ordinarily justifiable, when used by a dominant undertaking, *but* that certain circumstances can provide a justification (and therefore a *defence*) to the illegality of such a rebate. Nothing precludes an undertaking proving its justifiability, and it is submitted that this approach, whilst taking a harsh standpoint in respect of retroactive rebates, is a desirable development of the law, provided that objective justifications are taken seriously.

This section has explained the different ways in which dominant undertakings can show their rebates, though presumed to be illegal, to be justifiable on the facts. It has also been shown that the Commission is cognisant of such justifications. The goal has not been to explain how these justifications fit into the current binary structure of the defence – the prospect of additional objective justifications is clearly accepted by the CJEU in *Post Danmark I* (via the 'in particular' caveat), thus taxonomy is (if anything) merely a superficial issue. The issue is that, whilst the Guidance Paper and the jurisprudence of the Courts point to certain possible justifications, there are no examples of their successful pleading by a dominant undertaking. This does not mean that the objective justifications are virtually unavailable, but does show that, on the facts of the previously cited Decisions and cases, the Commission and the Courts did not deem a justification to have existed. This is not problematic for my argument – I have simply submitted that the qualified *per se* rule is justifiable if objective justifications are taken seriously: I have not claimed that previous adjudications regarding objective justifications have been mistaken (unfortunately, there is no space in this article for such a discussion). As such, my argument that objective justifications must be taken seriously is prospective, because we have not yet seen a factual scenario which yields these justifications.

[94] Pablo Ibáñez Colomo, 'AG Kokott in Post Danmark II: a legal test for quantity rebates' (*Chilling Competition*, May 2015) <https://chillingcompetition.com/2015/05/28/ag-kokott-in-post-danmark-ii-a-legal-test-for-quantity-rebates/> accessed 21 September 2016.
[95] See section 2(C) above.
[96] *British Airways* (n 19) [4].

5. *Remaining questions*

The CJEU's judgment in *Post Danmark II* had important impacts on the treatment of retroactive rebates, the relevance of the as-efficient-competitor test, and the requisite likelihood of exclusionary effect in order to find an Article 102 infringement. The primary interest in this article has been the impact on retroactive rebates, not just because *Post Danmark II* indicated a convergence in the CJEU's approach to rebates which can be seen as presumably abusive, but also because it is this aspect of the judgment which might be further developed in the near future. The *Intel* appeal will afford the CJEU an opportunity to deal with some unresolved issues which arise from *Post Danmark II*, either because they were not squarely addressed by the judgment, or because the effects of the judgment may be of interest in *Intel*.

An issue that was *not* squarely addressed by the CJEU in *Post Danmark II* is the three-fold classification which the General Court used in *Intel*, and which seems to be supported by prior case law. As it has been noted, Sidiropoulos had little doubt as to the confirmation of this classification,[97] but I have already suggested that the endorsement of the three-fold classification was equivocal. Furthermore, it is worth noting that AG Kokott's opinion claimed it to be 'ultimately immaterial whether the scheme can be assigned to a traditional category of rebate', with the important issue being 'whether the dominant undertaking grants rebates which are capable of producing on the relevant market an *exclusionary effect* which is not economically justified'.[98] The CJEU did not deal with this claim in *Post Danmark II*, but it is submitted that the qualified *per se* prohibition that flows from the case suggests an approach that treats retroactive rebates and exclusivity rebates in the same manner – as such, the CJEU may, in the *Intel* appeal, reconsider the three-fold classification, either by moving to a two-fold classification, or by placing exclusivity rebates and retroactive rebates in the same category and retaining a residual category. In light of the lack of clarity as to the CJEU's view on the three-fold classification, *Intel* will surely provide an inescapable opportunity for a clear view on the General Court's category-based approach to be given by the CJEU. It must be noted that the classification will *not* dictate the outcome in *Intel* – the significance of objective justifications postulated by this article is, it is submitted, a more important issue for the purposes of the outcome, given that the starting point is essentially of a presumption of illegality.

The rebates in *Intel* were generally exclusive,[99] thus the position of retroactive rebates is unlikely to be developed greatly from that which follows from *Post Danmark II*. However, it is possible that, depending on the CJEU's evaluation of the three-fold classification approach, the uniform approach recognized above might become more clear. That is, if the CJEU was to show itself to be more concerned with, as AG Kokott proposed, the likely effect of the rebates, rather than their categorisation, then it is possible that the judgment will shed light on the CJEU's views on exclusivity rebates in comparison to retroactive rebates.

6. *Rebates post*-Post Danmark II

The *Post Danmark II* judgment was significant in a number of ways, and this article has focused predominantly on one such way: the tightening of the approach to retroactive rebates, and indications of convergence with the treatment of exclusivity rebates. Undoubtedly, there remain differences between the two approaches, but this article has argued that the CJEU has recognized, in *Post Danmark II*, that retroactive rebates offered by a dominant undertaking will typically be anticompetitive, and that whilst an effects-analysis is allegedly necessary, the almost inevitable result of that analysis is that such a rebate scheme is an infringement of Article 102. In that respect, there is much less that separates retroactive rebates from exclusivity rebates, and retroactive rebates appear to

[97] Sidiropoulos (n 46).
[98] *Post Denmark II* (n 1), Opinion of AG Kokott, [29]
[99] This is subject to the minor exception of the rebates conditional upon 80% or 95% custom, which were still greatly conducive to fidelity.

be subject to a qualified *per se* rule. This results in a simpler and clearer approach, but is only justifiable insofar as objective justifications are sufficiently available as a defence and thoroughly considered by the Commission and the Courts. This article has sought to show how justifications for retroactive rebates can be demonstrated and should be recognized by the Commission.

Can International Law Manage Refugee Crises?

Azfer Ali Khan*

Introduction

This article shall aim to answer whether international law, as a legal framework, has the necessary tools to effectively deal with refugee crises in general. I shall use the current Syrian crisis as a springboard to contextualize some pertinent issues and determine whether, as some argue, international law has failed asylum seekers who seek refuge in States other than the home State. I will tackle the question in four parts. Part 1 will introduce the rights of a refugee and discuss the identification of a 'refugee' under international law. Part 2 will discuss the obligations of States under international law. Part 3 will then evaluate whether international law has had any substantial impact in protecting the rights of refugees, and if so, whether and to what extent it can be improved. Lastly, in Part 4, I will discuss a few proposed reforms and show why they are unlikely to take hold. With regards to the 'crises', the only focus of this article will be whether asylum seekers have substantial rights (which can be or are being protected) that can be invoked to offer them some form of protection against States in the international or regional context. This article is not concerned with how to manage crises in domestic contexts, or whether international law imposes a legal responsibility on States to take positive steps, such as by offering aid or offering protection through their armed forces.

The term 'manage' is understandably very broad. For the purpose of this article, I shall consider (a) whether international law guarantees some level of protection for refugees fleeing from their home state, (b) whether international law imposes a legal (as opposed to a political or moral) obligation on States to offer asylum to refugees and (c) whether there is empirical evidence to show that international law can be used to effectively protect the rights of refugees (i.e. whether there are instances where asylum seekers have successfully claimed municipal or international protection through legal process). I argue that the three questions should be answered in the affirmative, and if so, then one must conclude that international law has the potential to solve refugee crises, present and future. However, the fact that it is unable to do so, as evidenced by the chaos being endured by the Syrian asylum seekers (arguably as a result of a lack of systematic protection), suggests a pressing need for reform and points towards the conclusion that the international legal framework is in fact incapable of solving refugee crises as it stands today.

The meaning of key terms will now be clarified. International law refers to customary international law as well as multilateral treaty obligations, primarily those imposed by UN conventions. As a preliminary point, I use the term 'refugee' as a technical legal term to denote those who fulfil the definitional requirements of being a refugee under international law, whereas the term 'asylum seeker' shall refer to those whose status has not yet been determined. It is also to be noted that international refugee law is distinct from international human rights law (albeit with a significant degree of overlap), and this article is not specifically restricted to the former. The protection of asylum seekers under international human rights law will be taken into account when discussing whether the

* Magdalene College, Cambridge. This essay is dedicated to all my supervisors. In particular, to Dr Fernando Bordin for introducing me to the nuances of International Law. I wish to express my sincere gratitude to Mr Nigel Lim, Mr Matthew Wong and Ms Rebecca Halbach for their insightful comments. All errors remain my own.

international legal framework is potentially capable of presenting any meaningful solution to refugee crises. This article will also discuss relevant regional law, such as that of EU and AU.

Background: The Ongoing Crisis

An estimated nine million Syrians have fled their homes since the outbreak of civil war in March 2011. As of 21 December 2015, one million asylum seekers have stepped onto EU soil seeking safe haven from the violence in war-torn Syria. Affected countries are divided on how best to resettle these asylum seekers: some cite a legal basis for giving them protection, and others cite a political basis for turning them away.

The 9 million Syrians who have escaped include both those who have fled to their immediate neighbours and EU nations, and those who are internally displaced within Syria.[1] The first question I will try to answer will be in relation to the former: What rights does an individual asylum seeker have under international law? This will be my focus in the first half of Part 1.

There is evidence to show that most Gulf States in the region have passed various measures that aim to facilitate the entry and stay of Syrians since the outbreak of civil war,[2] and only about ten percent of the asylum seekers have fled to Europe.[3] My focus is particularly on the second of these types of asylum seekers, those who seek refuge in the broader international context rather than within Syria or in neighbouring States. The second question I will try to answer, therefore, will be in relation to whether these asylum seekers classify as refugees. This will be dealt with in Part 1B and subsequently in Parts 3 and 4.

Popular media has readily criticized the international community as a whole for their failure to provide refugees with shelter[4] and an adequate standard of living.[5] The third question I will try to answer will be whether States are under any legal obligations with regards asylum seekers or refugees. This will be answered in Part 2 and subsequently in Parts 3 and 4.

The fact that there has been so much trenchant criticism of the system suggests that the international community is simply not equipped to facilitate a surge of people seeking asylum. To answer whether this is a failure of international law or a failure of the international community in adhering to the law I first turn to analyse what the law actually is.

---

[1] Migration Policy Centre, 'Syrian Refugees' (Syrian Refugees, January 2013) <http://syrianrefugees.eu/> accessed March 23 2016.

[2] Gulf Labour Markets and Migration, 'A Note on Syrian Refugees in the Gulf: Attempting to Assess Data and Policies', GLMM - EN - No. 11/2015.

[3] World Vision, 'What You Need to Know: Crisis in Syria, Refugees, and the Impact on Children' (World Vision, March 9 2016) <http://www.worldvision.org/news-stories-videos/syria-war-refugee-crisis> accessed March 23 2016.

[4] BBC News – Middle East, 'BBC News, 'Syria Conflict: "Global Failure" to Protect Refugees' (BBC News, September 3 2015) <http://www.bbc.com/news/world-middle-east-34140403> accessed March 23 2016;
Rouba Al-Fattal, 'Europe may be failing Syrian refugees, but Canada shouldn't boast yet' (The Globe and Mail, March 17 2016) <http://www.theglobeandmail.com/opinion/europe-may-be-failing-syrian-refugees-but-canada-shouldnt-boast-yet/article29265806/> accessed March 21 2016; Esra Kaymak Avci, 'Syrian failure has "incalculable" costs: UN official' (AA, March 16 2016) <http://aa.com.tr/en/politics/syrian-failure-has-incalculable-costs-un-official/538112> accessed March 21 2016; Amnesty International, 'The world's pitiful response to Syria's refugee crisis' (Amnesty international, December 5 2015) <https://www.amnesty.org/en/latest/news/2014/12/world-s-pitiful-response-syria-s-refugee-crisis/> last accessed March 21 2016.

[5] France-Presse, Agence 'UN Syria Investigators Denounce "failure" to Protect Refugees' (GlobalPost, September 3 2015) <http://www.globalpost.com/article/6641821/2015/09/03/un-syria-investigators-denounce-failure-protect-refugees> accessed March 23, 2016.

1. *On Refugees*

A. What Rights Does a Refugee Have?

The notion of 'legal protection' has a very narrow focus.[6] The 1951 Convention relating to the Status of Refugees[7] (hereafter '**the 1951 Convention**') only applies once a State can identify the status of the party. This means that states are only bound by the Convention once an asylum seeker has been identified as a 'refugee' under the Convention. This may be problematic as it can insert a measure of discretion for the States, which may lead to States using procedure and technical rules to circumvent their international law obligations. In the 1951 Convention, States agreed to provide refugees with a wide range of basic rights, such as administrative assistance (Article 27), the grant of permission to transfer assets (Article 30) and the facilitation of naturalization (Article 34). Moreover, the 1951 Convention proposes as a minimum standard that refugees should receive 'at least the treatment which is accorded to non-citizens generally'.[8] The 2004 Qualification Directive,[9] which defines how the EU States define the notion of 'refugee', further calls for 'subsidiary protection' in the case of those who would face a real risk of serious harm if returned to their country of origin,[10] and this clearly applies to a large majority of those coming from Syria and the surrounding regions.

The cornerstone of the 1951 Convention is arguably the principle of non-refoulement, as provided for in Article 33. It states that all persons should enjoy the right not to be deported to a country where they may be subjected to persecution.[11] As this is almost certainly a principle of customary international law,[12] it should apply to all countries regardless of whether they are a party to the 1951 Convention. As such, it is quite clear that the 1951 Convention has codified some parts of customary international law, especially those relating to the basic rights of such refugees under articles 14 to 30.

Since they are considered a part of custom,[13] a breach of customary international law attributable to the State in question will invoke State responsibility for the internationally wrongful act under Article 1 of the 2001 ILC Articles on Responsibility of States for Internationally Wrongful Acts ('**ARS**').[14] The State will be required to make full reparation[15] for the breach under Article 31, and if the obligation is owed to the internationally community as a whole (i.e. an *erga omnes* obligation) under Article 48(1)(b), any State will be able to invoke the responsibility.[16] Given that

---

[6] Goodwin-Gill, Guy S, 'The International Law of Refugee Protection.' *Oxford Handbooks Online*, 36-*37.*

[7] 1951 Convention relating to the Status of Refugees (adopted 22 April 1954) 189 UNTS 137, in force for 147 states as of January 2016.

[8] Goodwin-Gill (n 6) 41.

[9] EC Directive 2004/83 of 29 April 2004, OJ L 304, 12.

[10] Goodwin-Gill (n 6) 42.

[11] For more information on this principle and what it entails, see Weissbrodt, David S, *The Human Rights of Non-citizens* (OUP 2008) Chapter 6.

[12] UN High Commissioner for Refugees (UNHCR), 'The Principle of Non-Refoulement as a Norm of Customary International Law. Response to the Questions Posed to UNHCR by the Federal Constitutional Court of the Federal Republic of Germany in Cases 2 BvR 1938/93, 2 BvR 1953/93, 2 BvR 1954/93' (UNHCR, 31 January 1994) <http://www.refworld.org/docid/437b6db64.html> accessed 23 March 2016.

[13] Hofmann, Löhr, Introduction to Chapter V. in A Zimmermann (ed), *The 1951 Convention Relating to the Status of the Refugees and its 1967 Protocol: A Commentary* (OUP 2011) 1089.

[14] International Law Commission, *Articles on Responsibility of States for Internationally Wrongful Acts*, November 2001, Supplement No. 10 (A/56/10), chp.IV.E.1.

[15] In addition to the ARS, see *Factory at Chorzow (Germany v Poland)* [1927] P.C.I.J. (ser. A) No. 9 for a discussion of the term 'reparation'.

[16] *Legal Consequences of the Construction of a Wall in the Occupied Palestinian Territory,* Advisory Opinion [2004] ICJ Rep 136, using the language of art 41 ARS. See *Barcelona Traction, Light and Power Company, Limited (Belgium v Spain)*, Second Phase, ICJ Rep 1970 3, 32 for further comments on the precise scope of *erga omnes* obligations.

refugee law overlaps with human rights law, and that the basic principles of non-refoulment and of temporary refuge are customary in nature, it is highly likely that such obligations will be *erga omnes.* The application of such obligations varies, since it depends on the political will of States to take positive steps to enforce measures. Although it is particularly effective in isolated situations, it has yet to apply in a meaningful manner if a large number of States are potentially in breach, as is the case in the current context. This differs from specific regional enforcement, such as an obligation to respect the rights of refugees articulated in the ECHR or 1969 Convention on the Organization of African Unity (hereafter '**the OAU Convention**').[17]

Therefore, it seems clear that once recognized as a refugee, a person has a significant array of rights that are protected under customary international law. This begs the question of why the rights of Syrian asylum seekers are not protected, as seems to be evident from the numerous media sources mentioned in the introduction. Some suggest that the answer lies in the way refugees are defined under the current legal framework.

B. Who is a Refugee?

Under international refugee law, Article 1(A)(2) of the 1951 Convention states that the term 'refugee' applies to any person who is outside the country of his nationality, owing to a well-founded fear of persecution for reasons of race, religion, nationality, membership of a particular social group or political opinion, and is consequently unable or unwilling to avail himself of the protection of that country. This is the 'centrepiece of international refugee protection' today.[18] It is further supplemented by subsidiary protection in regional bodies and through progressive development of international human rights law.[19]

Article (1)2 of the 1967 Protocol Relating to the Status of Refugees[20] (hereafter '**the Protocol**') amends the 1951 Convention and provides that the definition of the term 'refugee' is to be applied without the temporal and geographical limitations present in the text of the 1951 Convention. The definition of refugees under the Convention is generally seen as the minimum standard definition for the status of a person as a refugee[21] and forms the basis for the formulation used in the European Union.[22] Moreover, under international human rights law, Article 14 of the Universal Declaration of Human Rights states that '[e]veryone has the right to seek and to enjoy in other countries asylum from persecution'.[23] This applies to all asylum seekers and only seeks to supplement the definition in the 1951 Convention, thereby giving asylum seekers the protection of both human rights law and refugee law.

As for regional protection, the key (and perhaps the only practical) difference between the 2004 Qualification Directive and the 1951 Convention is the former's limitation that only nationals of non-EU countries and stateless persons are included in the scope of the Qualification Directive. Anja Klug suggests that the relevance of this limitation should not be overstated,[24] and that the relevant

---

[17] Organization of African Unity (OAU), *Convention Governing the Specific Aspects of Refugee Problems in Africa* (adopted 10 September 1969) 1001 U.N.T.S. 45.
[18] UNGA Res 429(V) (14 December 1950).
[19] For example the OAU (now AU) Convention 1969, the EC Directive 2004/83/EC, the Cartagena Declaration on Refugees 1984.
[20] 1967 Protocol Relating to the Status of Refugees (adopted 4 October 1967) 606 UNTS 267 art 1 para 1
[21] R Wolfrum (ed), *The Max Planck Encyclopedia of Public International Law* (OUP 2012).
[22] Council of the European Union Directive 2004/83/EC 2004 OJ L304/12, Ch. I General Provisions, art 2(c), on minimum standards for the qualification and status of third country nationals or stateless persons as refugees or as persons who otherwise need international protection and the content of the protection granted.
[23] Universal Declaration of Human Rights (adopted 10 December 1948) UNGA Res 217 A(III) (UDHR) at 71.
[24] Anja Klug, 'Harmonization of Asylum in the European Union: Emergence of an EU Refugee System?' 2004 GYIL 47, 594, 600.

procedure is still the one modelled on the 1951 Convention. This is especially true since EU nationals can bring a direct human rights claim to the ECtHR if their freedom of movement is obstructed. Moreover, as Hugo Storey points out, the Qualification Directive does not abrogate any of the provisions in the 1951 Convention, but only serves to add more detail to them.[25] As such, there is no substantial difference in the basic rights accorded to asylum seekers under EU law, though it is at least arguable that EU law may go further than customary international law in general by adding a level of subsidiary protection if the asylum seekers are fleeing from a real risk of serious harm in their home country.

Furthermore, as an extension of the 1951 Convention, the OAU (now AU) seems to encompass a slightly broader scope of protection.[26] The definition in the OAU Convention does not require a 'well-founded' fear of persecution for one to claim refugee status; rather, it opts for an objective examination of whether the facts of a specific situation fit within the definition's specified causes of flight. This is, at least arguably, significantly broader than the 1951 Convention and I shall discuss this further in Part 4 on how this definition might serve as a model to improve on the current general framework.

The main advocate of international protection for refugees is the Office of the UNHCR, which has a broad-ranging mandate with respect to refugees.[27] With regard to effectiveness, the general role of the UNHCR is not to provide direct assistance to refugees, but to coordinate the action of States along with private and public organizations, as well as to provide material assistance for 'persons of concern to the UNHCR'.[28]

2. On States

It is questionable whether States have any substantial obligation under customary international law to offer asylum to refugees. The 1951 Convention imposes a duty on States not to obstruct individuals' right to seek asylum,[29] rather than actually giving individuals the right to asylum itself.[30] The 1967 Declaration on Territorial Asylum affirms this in Article 1,[31] emphasizing that each State is the sole judge of the grounds upon which it will extend protection to asylum seekers. Article 2 tells us that other States 'shall consider' measures to lighten the burden when a State finds it difficult to manage asylum seekers. Since the 1967 Declaration, there has been little agreement within the international community with regard to reaching a universal instrument of asylum.[32] Instead, progress is generally seen in the form of regional agreements,[33] and even this has arguably been quite limited. For example, even in regional agreements like those under the European Convention of Human Rights (ECHR), a right to asylum has not been recognized.[34] The furthest that signatory States in the European region have gone is to recognize that the rights under both EU law and the ECHR stem from Article 14 of

[25] Hugo Storey, 'EU Refugee Qualification Directive: A Brave New World?' International Journal of Refugee Law (2008) 20 (1) 8, 9.
[26] Article 1(2) of the OAU Convention states that 'the term refugee shall also apply to every person who, owing to external aggression, occupation, foreign domination or events seriously disturbing public order in either part or the whole of his country of origin or nationality, is compelled to leave his place of habitual residence in order to seek refuge in another place outside his country of origin or nationality.'
[27] Statute of the Office of the United Nations High Commissioner for Refugees, GA res 428 (V), annex, 5 UN GAOR Supp (No 20) at 46, (1950) UN Doc A/1775 (Commonly referred to as '**the UNHCR Statute**').
[28] Weissbrodt (n 11) 164.
[29] Guy Goodwin-Gill, *The Refugee in International Law* (OUP 1998) 358.
[30] Elizabeth J Lentini, 'The Definition of Refugee in International Law: Proposals for the Future' (1985) 5 B.C. Third World L.J. 185.
[31] UN General Assembly, *Declaration on Territorial Asylum*, 14 December 1967, A/RES/2312(XXII).
[32] Goodwin-Gill (n 29) 365.
[33] ibid 366.
[34] ibid 366, 367-8.

the Universal Declaration of Human Rights (UDHR), which only crystallizes the right of non-refoulment and the right of individuals to have their cases considered on their merits.[35]

The AU seems to be encompass a slightly broader scope of protection,[36] as I have already stated. Further, inclusion of the phrase 'in either part or the whole of' the definition does away with the requirement that a person must first seek 'internal flight or relocation alternative' as required by 1951 Convention.[37] In effect, the differences led to the widening of the scope of those who would qualify for refugee status under the OAU Convention as compared to the 1951 Convention. However, this does not go so far as to recognize an international obligation to grant asylum to asylum seekers.[38] Ultimately, there remains insufficient State practice or *opinio juris* to demonstrate a duty of States to grant asylum to those seeking it under customary international law,[39] and judging by the uncertain and mixed treatment currently, the obligation to grant asylum to asylum seekers is unlikely to be incorporated as custom in the foreseeable future.

This is not to contradict the basis of the rights discussed in Part 1 above. Article 31 of the 1951 Convention provides support for a limited right of temporary admission for asylum seekers, even if they do not possess the requisite documents, to access fair and effective refugee status proceedings.[40] However, this requires evidence of a 'good cause' for entering the State in such a manner, and no doubt violence and imminent threats to life (as is the case in Syria) fall under this category.[41] This, combined with a customary duty of States to implement their treaty obligations in good faith under the 1969 Vienna Convention on the Law of Treaties,[42] will lead to an invocation of State responsibility if a State breaches its obligations. A State will be found to be acting in bad faith even if it tries to avoid its obligations.[43] The same analysis as in the previous subsection follows if a State is found in breach of its international obligations.

3. *On a Potential Solution*

In this section, I consider the arguments for and against the current international refugee law framework, and whether it is capable of offering any meaningful solution to refugee crises in general. The three criteria I will use are (a) the level of protection offered to asylum seekers, (b) the extent of the legal obligations on states for asylum seekers and (c) instances where asylum seekers have successfully asserted their rights as refugees. Almost necessarily, (c) will show how extensive (b) is.

A. The Level of Protection

The most pressing problem that international refugee law faces is that it suffers from a 'serious legal lacuna'[44] in that it determines the rights of those who are classified as refugees, but fails to bind States on how to carry out this determination.[45] Instead, it leaves States to rely on domestic laws or regional instruments to conclude whether a person qualifies for refugee status and whether, as a consequence,

---

[35] Goodwin-Gill (n 29) 368.
[36] n 26.
[37] 1951 Convention relating to the Status of Refugees (adopted 22 April 1954) 189 UNTS 137 art 1(A)(2).
[38] Goodwin-Gill (n 29) 368.
[39] Gregor Noll, *Negotiating Asylum* (Martinus Nijhoff 2000) 357–62.
[40] Goodwin-Gill (n 29) 384.
[41] ibid 387.
[42] 1969 Vienna Convention on the Law of Treaties (adopted 23 May 1969) 1155 *UNTS* 331, arts 26, 31.
[43] UNHCR, 'Written Case', 17 *IJRL* 427 (2005) 32.
[44] Hofmann, Löhr (n 13) 1087.
[45] ibid 624 'This means that most Third World refugees still remain literally excluded, as their flight is more often prompted by natural disasters, war, or broadly based economic turmoil than by 'persecution'. Against this background, it remains the task of a new worldwide instrument to adapt the 1951 Convention and the 1967 Protocol to the present needs'.

their substantive rights will be protected.[46] This clearly predicates the effectiveness of international refugee law on the effectiveness of municipal law, and arguably can be seen to undermine the international legal framework.

Two important caveats should be noted with regard to the definition of a refugee under the 1951 Convention that further labour this point. First, there is no universally accepted definition of 'persecution'. As such, Joan Fitzpatrick criticized that the 'elasticity of the definition of persecution depends upon the political will of States Parties implementing the Convention'.[47] I argue that ultimately, even with legal safeguards in place, there is wide room for States to circumvent their obligations, and hence, a State has a broad-ranging discretion whether or not to accord an asylum seeker the title of 'refugee'. Furthermore, it seems to be the case that most asylum seekers are in fact not facing a fear of persecution (such as those in Syria, for example), but instead have had their lives torn apart by war or natural disasters, and therefore are not considered refugees. This does not fall within the scope of the 1951 Convention and hence States can 'legitimately' refuse to offer protection to these asylum seekers.[48] Second, one can lose his or her status as a refugee if the circumstance in connection with which he was recognized as a refugee ceases to exist.[49] This point is less important in the Syrian context (since the conflict is ongoing), but nevertheless has a significant impact on asylum seekers' rights in general, especially when dealing with the long-term protection of refugees' rights. The issue of long-term protection ties in with the next criticism levelled against international law as a framework.

Some argue that international law, as a framework, is one that focuses on the migrant-refugee distinction,[50] and that the separation of the terms (a result of the 1951 Convention) has had a serious impact on 'long-term quality of protection' and the ability of refugees to access 'sustainable and meaningful solutions'.[51] While there is some merit in these arguments, I suggest that there is a meaningful difference between migrants and refugees, and that this goes beyond the mere distinction that a 'refugee *needs* admission whereas a migrant *wants* admission'.[52] The term 'refugee' indeed connotes a humanitarian form of aid, one which is, in many cases, only temporary, whereas the term 'migrant' generally implies a longer duration of residence, though not necessarily permanent. The difference therefore, is not simply with regard to the required aid at the point of entry, but also to the consequential (and hence, long-term) aid which stems from the difference between the two types of asylum seekers. Moreover, it is clear from the wording of the 1951 Convention[53] that the scope envisaged for refugees is primarily humanitarian in nature, in the sense that it offers a minimum standard of protection to those who cannot return home due to fear of persecution. It does not aim to provide a long-term or sustainable solution to the problem, although that would be ideal. The primary focus is to provide aid until the situation in the home territory is resolved. This lies in contradistinction to the status of a migrant, whose aim is usually to settle permanently in a new home. And while it is true that the two may overlap, and often do overlap, the distinction is an important one, especially in a crisis where there is a large influx of asylum seekers.

This is especially true from a pragmatic perspective. It is unlikely that all States will admit a universal duty to allow temporary entry for all 'migrants' (who do not classify as '**refugees**') since this inherently depends on the needs of the States as well as the skills of the migrants. It is easier to establish such a duty for refugees since the cause is at least primarily humanitarian in nature. This

[46] Hofmann, Löhr (n 13) 1088.
[47] Joan Fitzpatrick, 'Revitalizing the 1951 Convention' (1996) 9 Harvard Human Rights Journal 229, 240.
[48] Elizabeth J Lentini (n 30) 183.
[49] 1951 Convention (n 7) art 1(C)(5).
[50] Katy Long, 'When refugees stopped being migrants: Movement, labour and humanitarian protection', Migration Studies (2013) 1 (1) 4.
[51] ibid.
[52] ibid (author's emphasis).
[53] 1951 Convention (n 7) Preamble.

does not mean that there have been no negative ramifications as a result of the distinction;[54] it only means that there may be an argument for retaining the divide, especially in the context of a civil war crisis, such as the one in Syria. Since customary international refugee law does not include victims of natural disaster and violence or war, some may question how far we can classify customary international refugee law as humanitarian in nature. However, the important point is that compared to migrants, the cause is substantially more humanitarian. This is further evidenced by the fact that the term 'refugee', as it is used in ordinary language, includes all types of asylum seekers, and not just refugees under the technical legal definition. Thus, as I show below, when States take action to offer protection, it is explicitly couched in humanitarian terms – not legal terms.

Thus, the argument that the law has introduced an unnecessary division between asylum seekers (as between migrants and refugees) fails to convince. The criticism that it creates an unnecessary division between asylum seekers fleeing from persecution and those fleeing from war or disaster seems to hold much more persuasive force.[55] However, as noted above, international refugee law does place minimum obligations on States[56] that almost certainly apply to the context of Syria (and other similar large-scale crises). This is evident in the fact that virtually all States have recognized the refugee status of at least some of the Syrians seeking asylum.[57] Whether or not this recognition is due to legal reasons or political reasons, it follows that the obligations for States to give a basic degree of protection to those refugees can be invoked. To contextualize the point, over a 100,000 refugees have been granted asylum in Germany, with the rest of the EU countries trailing behind with numbers in the tens of thousands or less. The crux of the issue is not whether Syrian refugees are fleeing the country from a fear of persecution (for example the Yazidis or the Assyrian Christians fleeing from ISIS) or whether they are trying to escape the war-torn land (the large majority of the refugees). This is because States are not refusing them protection based on the asylum seekers' specific motivation,[58] but instead, offer protection based on their own interests. The reason why many countries take a stance one way or the other is because they either have political reasons to welcome refugees (like an ageing population in Germany) or political reasons to turn them way (like national security concerns and public backlash in the USA).[59] As such, the argument that determination of the status of asylum seekers is uncertain is not a strong criticism of the international community in the present context. The real issue is that technical refugee law has been overshadowed by major political concerns, and I shall discuss this further in Part 4.

Therefore, it is clear that there is a degree of protection for asylum seekers through international refugee law. The minimum standard definition in the 1951 Convention applies only to those who flee from persecution, but persecution as a term is not defined and the 1967 Declaration on Territorial Asylum clearly says that States are the sole judge of whether a person can be classified as a refugee or not.[60] Thus, we must conclude that there is a great potential for refugee law to offer a significant degree of protection to refugees, but not to asylum seekers, which is where the problem potentially lies.

---

[54] Alan Gamlen, 'An inborn restlessness: Migration and exile in a turbulent world' Migration Studies (2015) 3 (3) 307.
[55] Elizabeth J Lentini (n 30) 198.
[56] Alan Gamlen (n 54) 1089.
[57] Michael Martinez, 'Syrian Refugees: Which Countries Welcome Them' (CNN, September 10 2015) <http://edition.cnn.com/2015/09/09/world/welcome-syrian-refugees-countries/> accessed March 23 2016.
[58] Nina Shea, 'For Christians and Yazidis Fleeing Genocide, the Obama Administration Has No Rom at the Inn' (National Review, September 22 2015) <http://www.nationalreview.com/article/424401/christians-yazidis-persecuted-iraq-syria-refugees-excluded> accessed 11 April 2016; HL Deb 9 September 2015, col 1432-34.
[59] Rick Noack, 'This map helps explain why some European countries reject refugees, and others love them' (The Washington Post, September 8 2015) <https://www.washingtonpost.com/news/worldviews/wp/2015/09/08/this-map-helps-explain-why-some-european-countries-reject-refugees-and-others-love-them/> accessed 11 April 2016.
[60] Declaration on Territorial Asylum (n 36).

B. Imposing obligations on States

Another problem that arises with international refugee law is a lack of enforcement. Firstly, commentators argue that the law is not as effective as it could be. Since there is no dedicated UN agency to oversee refugees, the law receives a less than universal application.[61] Secondly, it is argued that there is an absence of an institutional mechanism that can directly enforce refugee rights, as international law, unlike domestic law, cannot force States to act.[62] However, it is clear that States regularly enforce these customary international law obligations on a municipal level.[63] Therefore, it is a full body of law that provides a framework in which States as legal entities interact, and to suggest otherwise would be pure rhetoric.[64] I have already shown how breaches of international obligations can invoke State responsibility and that this can be enforced through a variety of mechanisms.

The more relevant contention for the refugee crisis, therefore, is the first argument; namely, that there is no international agency that directly oversees the situation of refugees, since the UNHCR lacks authority to directly intervene with State discretion on whether or not to grant a person the status of a refugee. While this does mean that there is no international body that can evaluate whether States have fulfilled their obligations, this is not a fatal flaw in the machinery of refugee law. This is because there are a variety of transnational and regional bodies that are capable of overseeing refugee claims, and enforcement need not be directly implemented by UNHCR. The UNHCR usually acts as a mediatory that suggests guidelines and proposes solutions to tackle the difficulties resulting from a growing number of refugees.[65] For example, the European Commission has taken many steps to try to manage the flow of refugees into EU Member States.[66] Moreover, it is by no means a given that States would agree to such a body in practice, nor do I think it would it have an impact on the current refugee crisis due to the overwhelming number of people displaced. To prove this point, I shall turn to the last criterion, which focuses on specific instances where asylum seekers' rights have been affirmed.

C. Specific Instances of Rights' Protection

Without a doubt, the key milestone in international refugee law is the 1951 Convention. Although it does not prescribe a duty to accept refugees, it goes a long way in establishing the certainty of the rights which are protected, and the duties of States in respect of refugees, as I have already shown in the preceding two subsections. This is especially important in the international arena because, as Louise Holborn demonstrates, from the period of the First World War onwards, many States introduced municipal legislation giving the State the right to expel and return immigrants to their

---

[61] While there is the UNHCR, refugee law has no institution to evaluate state compliance, or to provide authoritative guidance on matters relating to refugee law. For more information see Hathaway and James 'Summary Conclusions of the Roundtable on the Future of Refugee Convention Supervision' (Opinio Juris, 27 March 2013) <http://opiniojuris.org/2013/03/27/summary-conclusions-of-the-roundtable-on-the-future-of-refugee-convention-supervision/> accessed March 23, 2016.

[62] Hersch Lauterpacht, 'Westlake and Present Day International Law' (1925) in Elihu Lauterpacht (ed.) *International Law: Being the Collected Papers of Hersch Lauterpacht*, vol 2 (CUP 1975) 401.

[63] While it would be impossible to cite an exhaustive list of cases, a great many can be found online. For example, the Center for Gender & Refugee Studies lists American cases, the European Database of Asylum Law lists EU cases and South African cases can be found at the Refugee Rights website (http://www.refugeerights.uct.ac.za/). See also UN High Commissioner for Refugees (UNHCR), *UNHCR Compilation of Case Law on Refugee Protection in International Law* (March 2008).

[64] James Crawford, *Chance, Order, Change: The Course of International Law, General Course on Public International Law* (Leiden Nijhoff 2013) Chapter 1.

[65] UN High Commissioner for Refugees (UNHCR), *The State of the World's Refugees: In Search of Solidarity* (2012) <http://www.refworld.org/docid/5100fec32.html> accessed 23 March 2016.

[66] European Commission 'Refugee Crisis: European Commission Reports on Progress in Implementation of Priority Actions' (Press Release, October 14 2015) <http://europa.eu/rapid/press-release_IP-15-5839_en.htm> accessed March 23, 2016.

countries of origin on the basis of discontentment.[67] This uncertain and unstable protection was insufficient to uphold the rights of refugees, and Hofmann and Löhr rightly argue that resettlement of these refugees thus became difficult since no country was seen to be legally responsible for individuals who could not receive protection from another State.[68] International refugee law, in tandem with regional legislative bodies and human rights law, now offers a much more certain bedrock of legal rules and offers States the opportunity to contribute to and develop the law over time, rather than be completely dependent on the political atmosphere of the day.

For example, through human rights claims, refugees have been able to assert their rights against States to force State action. Philip Leach gives an account of the decisions of ECtHR in *Sargsyan v Azerbaijan*[69] and the *Chiragov and Others v Armenia*[70] both of which upheld the rights of refugees against the States concerned.[71] He further suggests that this might put the Council of Europe in a position to pursue the interests of individual victims of the conflict. And though individuals have to assert their rights, it is still a level of legal protection that was not present before the implementation of the 1951 Convention. This is in contrast to purely political decisions, which, on the shaky foundations of policy, are almost always unpredictable.[72] More specifically in the context of the Syrian crisis, human rights law has been invoked to ensure the basic rights of refugees are upheld. This was the approach taken by the Lille Administrative Tribunal[73] in their decision to order France to improve the conditions of the Calais refugee camp.

Furthermore, the 2004 Qualification Directive, as I have already discussed, gives subsidiary protection that is wider than the 1951 Convention, and imposes a positive duty on EU states to offer protection to those who seek refuge from violence or war (and hence would apply to any 'crises'). Similarly, the OAU Convention gives a legal basis for protecting asylum seekers based on the context of the situation, and thus gives greater recognition to the need for humanitarian protection than the 1951 Convention.

Even in domestic law contexts, municipal courts have used international law as a basis to safeguard the rights of asylum seekers. For example, in the related Canadian cases of *R v Appulonappa*[74] and *B010 v Canada (Citizenship and Immigration)*,[75] the Canadian Supreme Court noted that Article 31 of the 1951 Convention meant that domestic law had to recognize that people would seek refuge in groups and work together to enter a State illegally. Therefore, Canada could not impose a criminal sanction on refugees solely because they aided others to enter illegally in their collective flight to safety. As such, it is clear that refugee law has shaped, and is in turn bolstered by, human rights law, regional law, and even domestic law.

The conclusion one must draw is that there is great potential from the mechanisms of international refugee law, as developed by the 1951 Convention, to offer a high degree of protection to refugees. This is primarily because customary international law has helped give form to national and municipal law.

---

[67] Holborn, *AJIL* 32 (1938), pp. 680, 683, 685; Weis, *BYIL* 30 (1953), pp. 478, 482.
[68] Hofmann, Löhr (n 13) 1091.
[69] App no 40167/06.
[70] App no 13216/05.
[71] Philip Leach, 'Thawing the Frozen Conflict? The European Court's Nagorno-Karabakh Judgments' (EJIL Talk, July 6 2015) <http://www.ejiltalk.org/thawing-the-frozen-conflict-the-european-courts-nagorno-karabakh-judgments> accessed March 23, 2016.
[72] David Martin 'YLS Sale Symposium: Interdiction of Asylum Seekers - The Realms of Policy and Law in Refugee Protection' (Opinio Juris, March 15 2014) <http://opiniojuris.org/2014/03/15/yls-sale-symposium-interdiction-asylum-seekers-realms-policy-law-refugee-protection/> accessed March 23, 2016.
[73] Lille Administrative Tribunal, 2 November 2015, No. 1508747.
[74] 2015 SCC 59.
[75] 2015 SCC 58.

4. *On Reform*

I have already shown that there is potential, through customary international law being developed by regional and municipal law, to effectively solve the three issues I posited in the introduction. However, in this section, I suggest that refugee crises will not be solved by international law and that the potential will remain just that: a potential, never an actuality. While this is clearly empirically true, given that the international community does not have a good track record in dealing with humanitarian crises in general (for example the Bosnian genocide, the Afghan civil war and the Rwandan genocide), what I suggest goes further: that international law is inherently incapable of utilizing the necessary tools (as they exist) to combat mass influx of asylum seekers to other regions.

A. The Main Problems

As Wessbrost notes,[76] there are three major issues that present a significant risk to what he terms the 'refugee regime': (a) the replacement of durable solutions with temporary protection; (b) the failure of States to share the burden of protecting refugees; and (c) the scaling back of refugee rights in response to the September 2001 attacks. These issues are underscored by the fact that the meaning of 'persecution' is still unclear, as States have not agreed on what persecution means, and the 1951 Convention definition does not offer protection for asylum seekers fleeing from natural disaster or violence.

The first issue has to do with what I have already discussed as a part of the general nature of the status of refugees, that the law as it stands today is ever only focussed on developing temporary rights for asylum seekers until the disaster they are escaping from ceases to exist. Anything further depends on political will, and that is and always will be uncertain.

The second issue relates to a lack of any duty under international law to displace asylum seekers from one country to another if the former lacks the capacity or infrastructure to deal with the influx. The impact of this issue has been reduced by regional organizations, because they offer an alternative venue to discuss how best to redistribute the asylum seekers. The clearest example where such discussions regularly occur is the EU and the Arab League. Although these discussions are political in nature, one has to recognize that there has been some progress in States trying to share the burden of protecting refugees.

The third issue is another way of defining political interests which curtail the effectiveness of the legal regime. National security, especially after the recent Brussels bombings, will inevitably play a key role in whether States choose to recognize asylum seekers as refugees and find themselves bound to offer protection or not.

These issues conclusively show that the issue of refugees is dominated by political concerns. The fact that there are a large variety of asylum seekers in crises, who may each have a different claim for protection, shows that it is increasingly difficult to sort out who should be a refugee and who should not. The problem becomes humanitarian in nature, and almost exclusively political. The fact is that States will be staunchly against any form of legal obligation imposed on them that derogates from their freedom to choose which asylum seekers to accept. Elizabeth Lentini gave a comprehensive account of the proposals made by some states to alter the definition of the 1951 Convention to include those who flee from violence and war, but the international community decisively rejected it.[77] I argue that states do not and will not accept to be bound to protect refugees in any further manner than the one already established in customary international law. For this reason alone, we must conclude that although there is potential to tackle the crises, the result will never be achieved. Some

[76] Weissbrodt (n 11) 178.
[77] Elizabeth J Lentini (n 30) 190-3.

authors suggest reforms that will encourage States to be more willing to protect the rights of asylum seekers, and I will show that each of the reforms is unlikely to bear fruit.

B. Three Possible, Alternative, Reforms

Most authors on international refugee law seem to agree that to approach these issues on a general level would be to attack the problem from the wrong direction, and that what we need is further decentralization to treat individual situations specifically in a regional setting.[78] In addition, I suggest that it may be valuable to treat smaller numbers of asylum seekers under a different legal framework, as compared to massive numbers of asylum seekers from crises like Syria.

The first reform is one that is region-centric. I am by no means the first to suggest greater regional protection. For example, an MPC report by Philippe Fargues and Christine Fandrich suggests the establishment of a Regional Protection Programme (RPP) to deal with the large influx of Syrian refugees.[79] The benefit of an RPP would be to provide sufficient assistance to the refugees on a regional scale, and it would be authoritative over member States, especially in times of crises. Further, it aims to 'continue positive asylum procedures throughout the EU, and grant prima facie recognition including provision of sufficient assistance to Syrian asylum seekers; encourage visa facilitation and family reunification for Syrians; and continue to work with its international partners to find a political and humanitarian solution to the Syrian crisis'.[80] Similar programmes could be introduced in other regional organizations as well, such as the AU. This would ensure (a) durable protection, (b) the sharing of the burden to protect refugees, and (c) a degree of certainty with regards to refugee rights.

However, through sustained, prolonged and divisive debates, States have shown that they are unlikely to agree to any global duty to protect refugees and give them shelter. Indeed, enforcing such a duty would be a logistical and practical nightmare. For example, on what (agreeable) grounds could we regulate the number of refugees a State could shelter? What would be the consequence of failing to protect them? A lack of consensus is evident from the inability of EU states to conclude any agreements on the issue of refugees, despite being a relatively unified economic region with a strong degree of authoritative central control from the Council as well as the different Courts. Further, I suggest that such a proposal would be best suited to a less volatile socio-political climate, especially not one mired in the midst of a global refugee crisis. Thus, although the scope for reform through this suggestion may be the only one that might have substantial effect in the long term, it is not practical as of now and hence we must consider other alternatives.

The second reform would be to subsume refugee law under human rights law. Some authors point out that the reach of human rights law is already impressive, and that human rights law already informs refugee law to a large extent.[81] While this is undoubtedly true, Colin Harvey rightly doubts whether conflating the two separate regimes would have any further benefit on the current state of the law. David James Cantor shows that the relationship between refugee law and human rights law is not necessarily consistent, as is generally assumed.[82] Moreover, separating the regimes of ordinary refugee applications from those driven by crises can be justified by legitimate resource constraints, which motivate the use of different criteria for refugee acceptance in different situations. This is not to say that States should be given liberty to turn a blind eye towards human rights or refugee rights, but

[78] For example, see Elizabeth J Lentini (n 30).
[79] Philippe Fargues, and Christine Fandrich, 'The European Response to the Syrian Refugee Crisis: What next?' Migration Policy Centre Research Report 2012/14.
[80] ibid 2.
[81] Colin Harvey, 'Time for Reform? Refugees, Asylum-seekers, and Protection Under International Human Rights' Law Refugee Survey Quarterly (2015) 34 (1) 43.
[82] David James Cantor, 'Reframing Relationships: Revisiting the Procedural Standards for Refugee Status Determination in Light of Recent Human Rights Treaty Body Jurisprudence', Refugee Survey Quarterly (2015).

rather that we should be cognizant of the practical limitations that each State faces, particularly where there is a mass influx of refugees. Lastly, there is no central body that adjudicates and enforces international human rights, as such, it seems slightly unhelpful to add international refugee law under the same category (other than the overlap which already exists).

The third reform would be to treat cases of mass influx of refugees under a different legal framework altogether. Ultimately, it is established that in cases of mass influxes of refugees, States already adopt different requirements to determine the status of an individual.[83] Some States may make it easier by granting prima facie refugee status to all, and others may make it more stringent by treating every case individually. In the name of preventing the ill-effects and human rights violations of the Syrian crisis from occurring again, it would be preferable to have a separate legal framework to deal with situations in which there is a mass influx of refugees. Situations with a large influx of asylum seekers should be managed on a regional level, instead of leaving it to State discretion. While the scope of this reform is unclear, I suggest that this reform, while attractive in theory, will be highly impractical. For one, the major issue is that States refuse to accept a universal obligation to accept and shelter asylum seekers. By introducing a different framework for a smaller number of asylum seekers, States would have the discretion to apply one of the other as they deem fit, unless specific rules are devised to make it an objective assessment. The question as to how this assessment will be formulated and enforced then needs to be answered. Moreover, this still avoids the problem that seems to be central: States have divergent interests in offering protection, and the only sustainable solution to refugee crises would be one that takes into account these differences.

5. *Conclusion: An Impossible Task*

As I have shown, the international framework which protects asylum seekers' rights in seeking refuge in other States is clearly not a failure. It provides minimum standards of protection, it is not always dependent on political will since domestic law, modelled on international law, can offer to protect the rights of some asylum seekers, and it has substantial effectiveness through regional and local authorities. The potential that exists in the legal framework however, is not being utilized. In September 2015, 674 international lawyers and practitioners signed an open letter criticizing the human rights violations being perpetrated against those seeking refuge.[84] There is a lot more that can be done.

I further suggest that it may not be viable to expect international law, as a legal framework, to be able to contribute further in a substantial way. This is because the international legal system, much like domestic legal systems, does not seek to police the individuals who draft it, but instead provides a set of agreements that are meant to be binding. Enforcement lies under the purview of the executive body, not the legislative body, and international law lacks a comparable enforcement system.

The current issue is one that is embedded far too deeply in the realm of politics. It is unlikely that international law, at least on a general level, will be able to compel States to act in a certain manner. However, international law can serve to publicize the plight of refugees and to provide a platform on which their rights can be discussed. It also offers an avenue of indirect, if not direct, legal protection. International refugee law is significantly bolstered by human rights law as well as regional and municipal law, but even then it lacks sufficient certainty and general application. This is why the international community as a whole, while it has been able to offer substantial help to asylum seekers, has also had some glaring failures. Thus, I conclude that international law will never be able to effectively 'manage' refugee crises.

---

[83] J-F Durieux and A Hurwitz, 'How Many is Too Many? African and European Legal Responses to Mass Influxes of Refugees' German Yearbook of International Law vol 47 (2004) 105.

[84] 'Open Letter to the Peoples of Europe, the European Union, EU Member States and Their Representatives on the Justice and Home Affairs Council' (Opinio Juris, September 22 2015) <http://ohrh.law.ox.ac.uk/wordpress/wp-content/uploads/2015/09/open-letter1.pdf> accessed March 23, 2016.

Considering the Best Interests Test in the Context of Disabilities

Vincent Ooi[*] and Jia Wei Loh[**]

Introduction

The United Kingdom is bound by several international obligations to eliminate discrimination against persons with disabilities, chief among these being the United Nations Convention on the Rights of Persons with Disabilities ('**CRPD**'), which was ratified on 8 June 2009. Compliance with these obligations is secured at the domestic level through provisions such as those in the Equality Act 2010 ('**EA 2010**'). However, parents with disabilities remain exceptionally vulnerable to losing the care and custody of their children under care orders and child arrangements orders. Thus, in 2006 the Social Care Institute for Excellence conducted a knowledge review which found that social workers and local authorities were, where these goals conflicted, less interested in keeping families with disabled members intact than in safeguarding children. These observations raise an important question: how should the law balance the best interests of children with the duty of the State to eliminate discrimination against persons with disabilities?

This article seeks to answer the question on the basis that the obligations owed by the State to parents with disabilities are distinct from those owed by parents to their children. It is imperative to map out the legal relationships between the parties to ensure that an attempt to address the issue of discrimination on the part of the State against parents with disabilities will not lead to a mistaken weakening of the obligations owed by parents to their children. In other words, it will be argued that an objective standard should apply to disabled parents as it does to able-bodied parents, because no child should have to accept a standard of care below a minimum threshold simply because one or both of his or her parents are disabled. With this important caveat in mind, the article will provide a critique of the current approach by domestic courts to deliberating (a) care orders under section 31 of the Children Act 1989 ('**CA 1989**') and (b) child arrangements orders under section 8 of the CA 1989 in the light of the welfare principle encapsulated in section 1 of the same Act. The domestic application of the welfare principle in relation to the parent-child relationship will be juxtaposed with the approach taken by the European Court of Human Rights ('**ECtHR**') in balancing the rights of the parent and child. It will be argued that the ECtHR's approach is preferable to the welfare principle adopted in domestic courts because the former better captures conventional understanding of the parent-child relationship. On that footing, the article will recommend the replacement of the welfare principle with Choudhry and Fenwick's 'parallel analysis',[1] which is modelled on the ECtHR's approach. As an alternative to the parallel analysis, the article will also demonstrate that Eekelaar's 'modified least detrimental alternative' analysis[2] is preferable to the welfare principle. The adoption of either of these alternatives would have an incidental effect in helping the State better address its obligations to parents with disabilities but not at an unacceptable cost to the child's welfare.

---

[*] Trinity College, Oxford
[**] Keble College, Oxford

This article is an extended version of a similarly-titled essay which won the second prize in the 42 Bedford Row Disability Law Essay Prize Competition 2016. The authors are grateful to 42 Bedford Row and the Organising Committee for their efforts in organising the Competition.

[1] Shazia Choudhry and Helen Fenwick, 'Taking the Rights of Parents and Children Seriously: Confronting the Welfare Principle under the Human Rights Act' (2005) 25(3) OJLS 453.

[2] John Eekelaar, 'Beyond the Welfare Principle' [2002] CFLQ 237.

In the upshot, the article will argue that the most sensible approach to balancing the best interests of children with the State's obligation to eliminate discrimination against parents with disabilities would be for the State to equalise the positions of disabled parents and their able-bodied counterparts under an objective test. Apart from jettisoning the welfare principle for either the 'parallel analysis' or the 'modified least detrimental alternative' analysis, several other legal and non-legal solutions will be proposed to address direct and indirect discrimination (such as subconscious bias) against parents with disabilities in the public law and private law proceedings on care orders and child arrangements orders respectively.

1. *The Duty to Eliminate Discrimination*

It is clear that the State has an obligation to eliminate discrimination against parents with disabilities, and that to separate a child from his or her parents on the basis of a disability of either or both of the parents, without more, would amount to discrimination. Article 23(4) of the CRPD specifically provides:

> States Parties shall ensure that a child shall not be separated from his or her parents against their will, except when competent authorities subject to judicial review determine, in accordance with applicable law and procedures, that such separation is necessary for the best interests of the child. In no case shall a child be separated from parents on the basis of a disability of either the child or one or both of the parents.

In addition, the State is bound by sweeping anti-discrimination provisions in several international instruments. For example, Article 26 of the International Covenant on Civil and Political Rights ('**ICCPR**') provides:

> All persons are equal before the law and are entitled without any discrimination to the equal protection of the law. In this respect, the law shall prohibit any discrimination and guarantee to all persons equal and effective protection against discrimination on any ground…

Another key source of obligations for the State is the Convention for the Protection of Human Rights and Fundamental Freedoms ('**ECHR**'), as incorporated into domestic law by the Human Rights Act 1998 ('**HRA 1998**'). The relevant article here is Article 8, which guarantees the right to respect for private and family life. Article 8 provides a substantive right for the parasitic Article 14 to latch on to, binding the State to ensure the prohibition of discrimination in cases where the right to respect for private and family life is engaged.

Article 8: Right to respect for private and family life

> Everyone has the right to respect for his private and family life, his home and his correspondence.

Article 14: Prohibition of discrimination

> The enjoyment of the rights and freedoms set forth in this Convention shall be secured without discrimination on any ground such as sex, race, colour, language, religion, political or other opinion, national or social origin, association with a national minority, property, birth or other status.

While Article 14 does not expressly prohibit discrimination on the basis of disability, Herring has suggested that there is good reason to think that the miscellaneous provision for discrimination based on 'other status' is wide enough to catch disability based on discrimination.[3] Indeed, one would

[3] Jonathan Herring, *Family Law* (7th edn, Pearson 2015) 545.

be hard-pressed to come up with principled reasons why discrimination based on disability would be permitted under the ECHR.

At this juncture, it is necessary to identify two key differences between, on the one hand, the ECHR, and, on the other, the CRPD and ICCPR. First, unlike the CRPD and the ICCPR, the ECHR allows for a dialogic model between the English courts and the relevant international judicial institution – in this case, the ECtHR. This model enables us to compare the two lines of jurisprudence in relation to the parent-child relationship, as will be set out later in the article. For one, section 2 of the HRA 1998 provides that a court or tribunal 'determining a question that has arisen in connection with a Convention right must take into account' any relevant Strasbourg jurisprudence. Further, the ECtHR has constantly reiterated the fact that national legislatures and courts are given a margin of appreciation as they 'are in principle better placed than an international court to evaluate the local needs and conditions and to decide on the nature and scope of the measures necessary to meet those needs'.[4] Secondly, the ECHR differs from the CRPD and ICCPR in terms of legal status. The UK adopts a dualist legal framework so that international obligations acquire legal force only when they are given effect to in the form of statutory provisions. While the HRA 1998 expressly provides a mechanism for the provisions of the ECHR to be applied in the domestic legal system (through sections 2, 3, 4 and 6),[5] there are no similar enforcement mechanisms for the CRPD and the ICCPR

At the domestic level, section 15 of the EA 2010 addresses direct discrimination arising from disability and stipulates that:

> (1) A person (A) discriminates against a disabled person (B) if—
>
> (a) A treats B unfavourably because of something arising in consequence of B's disability, and
> (b) A cannot show that the treatment is a proportionate means of achieving a legitimate aim.
>
> (2) Subsection (1) does not apply if A shows that A did not know, and could not reasonably have been expected to know, that B had the disability.

Further, section 19 of the EA 2010 prevents indirect discrimination by prohibiting the application of a provision, criterion or practice that is disproportionately discriminatory in relation to a relevant protected characteristic, which includes an individual's disability.[6]

The cumulative effect of these international obligations and domestic legislation, as Lady Hale noted in *Re B (A Child) (Care Proceedings: Threshold Criteria)*, is that 'the State does not and cannot take away the children of all the people who (…) suffer from physical or mental illnesses or disabilities'.[7] However, it is questionable whether all possible measures have been taken to ensure that parents with disabilities are not prejudiced against on the ground of their conditions.

---

[4] *Animal Defenders International v UK* (2013) 57 EHRR 607.

[5] As mentioned above, section 2 of the Act obliges the UK courts to take into account judgments and decisions of the ECtHR when deciding a matter that involves a Convention right. Section 3 requires the courts to construe legislation so that it is compatible with the Convention rights where it is possible to do so. Section 4 permits the courts to make a declaration that legislation is incompatible with the Convention rights. Section 6 makes it unlawful for a public authority to act in a way which is incompatible with a Convention right. A court is a public authority for these purposes.

[6] EA 2010, s 19(3).

[7] [2013] UKSC 33, [2013] 1 WLR 1911 [143].

2. *Direct and Indirect Discrimination (Including Subconscious Bias)*

Theoretically, discrimination can be divided into direct and indirect discrimination. The EA 2010 defines direct discrimination in section 13 as where a person treats someone less favourably than he would treat others because of a protected characteristic. Indirect discrimination is defined in section 19, which prohibits practices that disproportionately put a disabled person at a disadvantage. To put it simply, direct discrimination occurs where one is discriminated against merely on the basis of certain characteristics, while indirect discrimination occurs where a uniform standard is applied, but one unreasonably fails to vary the standard to allow for certain characteristics.

While the State is under an obligation to eliminate both direct and indirect discrimination, the two are treated quite differently. There is an absolute prohibition on direct discrimination, since there are arguably no grounds for treating people less favourably simply because they have certain protected characteristics. On the other hand, the law will not generally intervene where an objective, uniform standard is applied. In these cases, the discrimination may be justified as a proportionate means to a legitimate aim pursuant to section 19(2)(d) of the EA 2010. Where there is a legitimate difference in treatment warranted by certain characteristics, states tend to provide the bulk of assistance out of public funds and simultaneously require reasonable adjustments to be made to accommodate the relevant persons in their interactions with them.[8]

In the context at hand, the prohibition on direct discrimination is manifested in the requirement that social workers assess each parent objectively on his or her capacity to care for their children and not decide the case merely on the basis of the parent(s)' disability. It is unlikely that a social worker would consciously discriminate against parents simply because they were disabled. However, there may be subconscious bias on the part of social workers and local authorities where, due to an inadequate understanding of the abilities of persons with particular disabilities, they understate the competence of the parent and effectively discriminate against him or her. The article will propose solutions to address subconscious bias in due course. Further, the State has a duty to provide disabled parents with sufficient support so that their children receive at least the minimum standard of care. Its failure to discharge the duty would lead to unacceptable outcomes: (1) children whose parents have disabilities might have to accept a standard of care which falls below the minimum threshold; and (2) parents with disabilities would face indirect discrimination in relation to their right to care for their children.

3. *The Best Interests Test*

A. The Welfare Principle at English Law

The State's obligation to eliminate discrimination against parents with disabilities is distinct from the obligation to act in the best interests of children. The law as it stands safeguards the best interests of the child through the welfare principle set out in section 1 of the CA 1989, and the principle undergirds both the public law and private law proceedings discussed below. Subsection (1) provides:

> When a court determines any question with respect to—
>
> > (a) the upbringing of a child; or
> >
> > (b) the administration of a child's property or the application of any income arising from it,
>
> the child's welfare shall be the court's paramount consideration.

[8] Arlene S Kanter, *The Development of Disability Rights Under International Law: From Charity to Human Rights* (2015 Routledge), 38.

While much has been said about the welfare principle,[9] *J v C*[10] remains the leading case on its meaning and application. In that case, Lord MacDermott, with whom Lord Pearson expressly agreed, explained that the phrase 'paramount consideration' in section 1 of the Guardianship of Infants Act 1925 means:

> more than that the child's welfare is to be treated as the top item in a list of items relevant to the matter in question. [The words] connote a process whereby, when all the relevant facts, relationships, claims and wishes of parents, risks, choices and other circumstances are taken into account and weighed, the course to be followed will be that which is most in the interests of the child…[11]

This suggests that the child's welfare is ultimately the court's sole consideration, with other secondary factors being taken into account in order to determine the course of action that would maximise the child's welfare. That there is no room for considerations of parental rights per se was affirmed at the highest level in *Re G (Children) (Residence: Same-Sex Partner)*[12] and *Re B (A Child) (Residence: Biological Parent).*[13] In the former case, Baroness Hale, with the agreement of the rest of the Appellate Committee, reiterated that there 'is no question of a parental right'[14] and that the welfare of the child determines the course to be followed. In the latter case, Lord Kerr, delivering the leading opinion in the Supreme Court, emphasised that '*In re G* had given the final quietus to the notion that parental rights have any part to play in the assessment of where the best interests of a child lay',[15] and that '[i]t is only as a contributor to the child's welfare that parenthood assumes any significance'.[16] The uncompromising stance of the court in the application of the welfare principle means that parents with disabilities are expected to meet the best interests of the child to the same extent as their able-bodied counterparts.

While the CA 1989 did not define 'welfare', it has added flesh to the concept through the statutory checklist in section 1(3). The checklist is mandatory when the court is considering whether, *inter alia,* to make a section 8 child arrangements order or a section 31 care order.[17] For the purposes of this article, paragraph (f) of the subsection is especially significant because it takes into account the capability of each parent in meeting the child's needs. As Herring has pointed out,[18] the disability of a parent could have implications here. Thus, a disability could impede a parent from exercising his or her parental responsibilities, and render him or her particularly susceptible to losing the care or custody of the child.

B. Applications of the Welfare Principle at English Law

I. State-Parents Situations: Care Orders under CA 1989, s 31

When a local authority or the National Society for the Prevention of Cruelty to Children ('**NSPCC**') makes an application for a care order pursuant to section 31 of the CA 1989, it will often be placed in an antagonistic relationship with the parents of the child in question. Section 31(2) of the Act of 1989 stipulates the 'threshold criteria' for making a care order:

---

[9] See, for example, Helen Reece, 'The Paramountcy Principle: Consensus or Construct?' (1996) 49 Current Legal Problems 267.
[10] [1970] AC 668.
[11] *J v C* (n 10), 710-711.
[12] [2006] UKHL 43, [2006] 4 All ER 241.
[13] [2009] UKSC 5, [2009] 1 WLR 2496.
[14] *Re G* (n 12) [31].
[15] *Re B* (n 13) [33].
[16] ibid [37]. For a recent case following this line of authority, see *Re E-R (a child)* [2015] 2 FCR 385.
[17] CA 1989, s 1(4).
[18] Herring (n 3), 545.

A court may only make a care order or supervision order if it is satisfied—

(a) that the child concerned is suffering, or is likely to suffer, significant harm; and

(b) that the harm, or likelihood of harm, is attributable to—

(i) the care given to the child, or likely to be given to him if the order were not made, not being what it would be reasonable to expect a parent to give to him; or

(ii) the child's being beyond parental control.

At this juncture, it is important to note that the threshold criteria constitute an objective standard which does not take into account the characteristics or capabilities of the parents, including whether or not one or both of them have disabilities. As Hughes LJ observed pointedly in *Re D (A Child) (Care Order: Evidence)*:

> [If the requisite care] were to be judged by the standards of the parent with the characteristics of the particular parent in question, the protection afforded to children would be very limited indeed, if not entirely illusory. It would in effect then be limited to protection against the parent who was fully able to provide proper care but either chose not to do so or neglected through fault to do so.[19]

The objective nature of the threshold criteria can be readily justified. For example, in *X v Liverpool City Council*,[20] a care order was obtained after a parent had repeatedly, and despite warnings, driven in a car with the children even though he was legally classified as blind. It is clear that the rationale of the decision was to protect the children in question rather than to effect impermissible social engineering.[21] In the same vein, the court in *Re D (A Child) (Care Order: Evidence)*[22] observed that it is not discriminatory to assess a parent with learning disability by the standards demanded of an able-bodied parent.

It may be asked whether the threshold criteria should be applied in light of the provision of available support to a parent.[23] Hedley J's speech in the different context of a parent's care of a severely disabled child may be instructive here. He stated that 'a parent cannot be said to be responsible for a falling below the standard of "reasonable care" if the public authorities cannot or do not provide what would be reasonably necessary to support that parent'.[24] This could mean that the threshold criteria should be adjusted as against a parent with a disability who ought to have been provided with a reasonable level of support from public authorities. While this adjustment could potentially render disabled parents less susceptible to losing their child pursuant to a care order, it is also undesirable as it could open the floodgates to greater subjectivity in the threshold criteria at the expense of the child's welfare. Other solutions would be proposed in the course of the article.

II. Parent-Parent Situations: Child Arrangements Orders under CA 1989, s 8

Section 8 child arrangements orders arise in private disputes concerning children. They are broadly classified into residence orders, contact orders, specific issue orders, prohibited steps orders, and orders varying or discharging such orders.[25] In contrast to a section 31 care order, a section 8 child arrangements order may turn on the antagonistic relationship between a disabled parent and his or her

[19] [2010] EWCA Civ 1000, [2011] 1 FLR 447 [35].
[20] [2005] EWCA Civ 1173, [2006] 1 WLR 375.
[21] *Re SB (Children) (Care Proceedings: Standard of Proof)* [2009] UKSC 17, [2010] 1 AC 678 [17].
[22] [2010] EWCA Civ 1000, [2011] 1 FLR 447.
[23] Stephen Gilmore and Lisa Glennon, *Hayes and Williams' Family Law* (3rd edn, OUP 2012) 643.
[24] *LBH (A Local Authority) v KJ* [2007] EWHC 2798 (Fam) [22].
[25] CA 1989, ss 8(1) and (2).

able-bodied counterpart. Local authorities drop out from the picture in these private law proceedings[26] because it has been resolved that they can only seek to obtain parental responsibility for a child in care proceedings.

As in care proceedings, the court must consider the checklist of factors in the CA 1989, s 1(3).[27] It seems that the manner in which the checklist is used differs from one judge to another,[28] but it is abundantly clear that a decision would be liable to be overturned on appeal if a judge failed to consider an item on the checklist which was pertinent to the case.[29] The disability of one parent is likely to be relevant under the CA 1989, s 1(3)(f), where the court considers how capable each parent is of meeting the child's needs. In *M v M (Parental Responsibility)*[30] a father suffered from learning disabilities, which were aggravated by a motorcycle accident, so that he had defective reasoning powers and a mercurial temper. Wilson J decided that, notwithstanding the father's commitment to the child and the important relationship between them, it would be inappropriate to give the father parental responsibility because he was incapable of exercising the rights and responsibilities of parenthood. This and other similar cases demonstrate the need to explore ways to level the playing field between a disabled parent and a non-disabled one, so that the former's disability does not unfairly render him or her at a disadvantage in seeking parental responsibility over the child.

C. Problems with the Welfare Principle

While the English courts have been adamant that there 'is no question of a parental right',[31] that approach seems fundamentally at odds with the wording of the ECHR. Article 8 expressly provides that 'everyone has the right to respect for his private and family life'. It is manifestly clear that an application to remove a child from his parent *prima facie* entails a serious threat to family life and potentially violates that parent's Article 8 right. With respect, the UKSC cannot simply hold that there 'is no question of a parental right'. While the domestic courts may be reluctant to whittle down the protection afforded to children, it seems untenable to argue that parental rights are irrelevant or, worse, non-existent in this context. Further, as Choudhry and Fenwick noted, the Convention requires a 'careful and sensitive balancing act between competing rights and interests' and not 'a brief and mechanical recital' that interference with parental rights is a 'proportionate response'.[32] The 'failure to consider the Article 8(1) rights themselves prior to considering the consequentialist arguments under Article 8(2) amounts to a violation of the court's duty under s 6(1), HRA'.[33]

The current approach of the domestic courts also seems questionable when one considers the text of section 1(1) of the CA 1989, which provides that the welfare of the child shall be the 'paramount' consideration. With respect, it is most unusual for the courts to interpret 'paramount' as 'sole'. While the word 'sole' suggests the irrelevance of considerations apart from the welfare of the child, 'paramount' may more naturally be taken to mean that relatively greater weight should be ascribed to the welfare of the child vis-à-vis other pertinent considerations. While both interpretations might result in the same outcome in the vast majority of cases, particularly if the weight given to the welfare of the child greatly exceeds the weight given to all other factors, there is a significant theoretical difference between the two. To illustrate the point, it is not the same to say that: (a) parental rights are almost always overridden by considerations as to the welfare of the child, and (b) there are no parental rights at all. Given the foregoing analysis of the wording in the ECHR and the CA 1989, it is submitted that parental rights ought to be taken into account in decisions

[26] CA 1989, s 9(2).
[27] See n 17.
[28] Herring (n 3) 523.
[29] *Re H (A Child) (Contact: Welfare Checklist)* [2010] EWCA Civ 448, [2010] 2 FLR 866.
[30] [1999] 2 FLR 737.
[31] *Re G* (n 12) [31].
[32] Choudhry and Fenwick (n 1) 465.
[33] ibid.

involving the parent-child relationship, such as care orders and child arrangements orders, as will be discussed below.

Apart from arguments based on interpretation of the ECHR and the CA 1989, it seems contrary to conventional understanding to say that the welfare of the child is the sole consideration in circumstances turning on the parent-child relationship. Consider a case of a newborn infant with parents without disabilities, and suppose that, even after taking into account the blood relation with his natural parents, he would enjoy better welfare if he were brought up by a different set of parents. If the child's welfare were the sole consideration, it would be justifiable to remove him from his natural parents so as to be brought up by the other set of parents. The fact that this example would be unpalatable to right-thinking individuals suggests that parents have rights not to have their children removed from them so long as they can provide a minimum standard of care for them. This seems to have been considered by Strasbourg jurisprudence, which does recognize parental rights, and it seems puzzling why the domestic courts have been reluctant to recognise this. With respect, it is submitted that it is imperative to take into account parental rights in arriving at decisions with a bearing on the parents-child relationship.

D. The Best Interests Test under the ECHR

In comparison to the approach of the English courts, the European Commission of Human Rights has adopted an approach that seems more in line with the text of the ECHR. In *A and Byrne and Twenty-Twenty Television v UK*[34] the Commission squarely confronted the tension between the welfare principle and a parent's Article 8 right to respect for her family life. The first applicant claimed that her Article 8 right was violated as a result of the refusal of the domestic courts, upon application of the welfare principle, to accept her decision that her child should participate in a television programme about her development and education in an institution catering for her special needs. The Commission held that the national authorities must strike a fair balance between the relevant competing interests. In other words, a child's welfare is not the sole consideration and does not automatically prevail. This can be contrasted with the approach taken by the English courts.

E. Reconciling the Two Lines of Jurisprudence

Valiant attempts have been made by the English courts to reconcile the jurisprudence of the ECtHR and the English courts. The key idea seems to be that this is merely an issue of semantics, and that in actual fact the tests applied really are the same. In *Re KD (A Minor) (Ward: Termination of Access)* Lord Oliver observed that the conflict was only semantic and 'lies only in differing ways of giving expression to the single common concept'.[35] Further, in *Re B (A Minor)* Lord Nicholls opined that '[a]lthough the phraseology is different, the criteria to be applied in deciding whether an adoption order is justified under Article 8(2) lead to the same result as the conventional tests applied by English law'.[36]

While their Lordships may be correct insofar as the two tests tend to produce the same result in practice, it is equally clear that the two lines of jurisprudence differ considerably in terms of theoretical foundations. In particular, the Strasbourg line of reasoning furnishes a clearer theoretical framework, which is more helpful in practice. As Choudhry and Fenwick put it, the CA 1989 is 'rule-utilitarian' because the course of action taken is determined by the rule that the child's welfare should be paramount.[37] In contrast, the HRA 1998 exhibits 'qualified deontology' because it is rights-based although some of the rights, such as Article 8, are qualified by exceptions.[38] This theoretical difference

[34] 1998] 25 EHRR CD 159.
[35] [1998] AC 806, 825.
[36] [2002] 1 FLR 196 [31].
[37] Choudhry and Fenwick (n 1) 457-458.
[38] ibid.

has practical implications for the way the two courts approach family law cases. Thus, as Choudhry and Fenwick explained, 'in the family law sphere the approach of the European Court of Human Rights differs considerably from that of the UK courts since it seeks to balance opposing rights and does not start from the assumption that the paramountcy principle will determine the issue'.[39] It is only on the ECtHR's approach that one can directly weigh parental rights with the welfare of the child in making decisions bearing on the parents-child relationship.

4. *The Rights Between the Parties*

As established earlier, the State is obliged to ensure that the individual is not discriminated against as a result of his or her disability ('**State-Parents Obligation**'). In the particular context at hand, it is the State's duty to ensure that a disabled parent is not placed in a disadvantageous position by virtue of his or her disability when considering his or her right to family life. The State-Parent Obligation must be distinguished from the rights and duties between a parent and his or her child under the Best Interests Test ('**Parents-Child Obligation**'). The latter obligation should not be affected by third party interests outside the bilateral relationship between parents and their child. Returning to our earlier example in the article, it does not matter how much better a third party potential fosterer may be at parenting as compared with the child's parents. As long as the child is adequately cared for, it would not generally be acceptable to remove him from his parents.

The approach to decisions bearing on parents-child relationships adopted by the ECtHR is preferable to the welfare principle adopted in the domestic courts because it better reflects the parents-child relationship and gives us a better understanding of its nature. This in turn enables us to better distinguish the Parents-Child Obligation from the State-Parents Obligation. When looking at a complex bundle of obligations, one should begin by attempting to understand the nature of the relationships between the parties involved, as this will help identify the nature of the various obligations. On the ECtHR's approach, one recognises the existence of parental rights. Where these rights are incompatible with the rights of the child, we can weigh them against each other to determine the correct balance to be struck between these two rights. Where one or more parents have disabilities, it is clear on this model that it is not a consideration that affects the Parents-Child Obligation, because parents with disabilities owe their children exactly the same obligation as able-bodied parents. The State-Parent Obligation of the State to ensure that parents with disabilities are not at an unfair disadvantage due to their disability is external to this Parents-Child Obligation since there is no direct obligation between the child and the State in these two obligations.

In contrast, on the welfare principle, any interests that parents might have are merely considered as part of an assessment as to what would be the course of action to maximise the welfare of the child. It is unclear whether and, if so, how the State ought to intervene to ensure that parents with disabilities are not discriminated against in court proceedings and decisions which would potentially bear on their relationship with their children. Without being able to clearly map out the various obligations, there is a danger of a child being asked to shoulder some of the burden of a State-Parent Obligation, to which he or she was never supposed to be party to in the first place.

[39] Choudhry and Fenwick (n 1) 454.

5. *Legal Solutions*

The following section explores various legal solutions to eliminate discrimination against parents with disabilities in the contexts of care orders and child arrangements orders. To recapitulate the point, it is in the nature of the State-Parents Obligation that these solutions have to focus on supporting the disabled parent, and not the content of the Parents- Child Obligation within the Best Interests Test.

A. Alternative Conceptions of the 'Best Interests Test'

The foregoing analysis has suggested that parents with disabilities are expected to meet the requisite standard of care for their children to the same extent as their able-bodied counterparts. At the very least, a child should not have to compromise his or her interests beyond a minimum threshold merely because one or both of his or her parents have a disability. However, there are more nuanced conceptions of the 'Best Interests Test' that satisfies this requirement.

Eekelaar, for one, has proposed the 'modified least detrimental alternative' approach to balance the interests of the parent and the child.[40] He argues that the virtue of the welfare principle resides in the fact that 'it requires a decision made with respect to a child to be justified from the point of view of the child's interests', but makes the basic premise that '[a] decision need not be justified only from that point of view'[41] as long as the different interests at stake are clearly delineated and distinguished. There are several prongs to the approach: (a) the best course of action is that which avoids inflicting the most damage on the well-being of any interested individual, that is, his or her prospects of realizing significant life goals;[42] (b) it is easier to register a detriment to a child because of their general vulnerability and the potential longer-term effects on them; (c) no solution should be adopted where the detriments outweigh the benefits for the child, unless that would be the inevitable outcome regardless of the course of action taken;[43] and (d) there is a degree of detriment to which a child should never be subjected.[44]

It is interesting to note that the modified least detrimental alternative analysis can produce a radically different outcome in favour of a disabled parent. Eekelaar provides the following illustration. Suppose a parent (X), who suffers from a disability, seeks to keep the child (C) after separating from the other parent (Y). The loss of C would be more detrimental to X than to Y because of the assistance C can provide X. However, to remain with X would be a greater loss to C because of reduced opportunities and other restrictions arising from the need to help X. On an application of the welfare principle, C will go to Y because he or she will enjoy greater welfare with that parent. In contrast, on the modified least detrimental alternative analysis, C will *prima facie* go to X, unless that solution would cause an overall detriment to C.

Another alternative approach to balancing the interests of the parent and the child is Choudhry and Fenwick's 'parallel analysis', which is grounded in the ECHR.[45] One facet of this analysis is that, unlike the welfare principle, the best interests of the child do not prevail automatically. Rather, there must be a detailed consideration of all the parties' rights and interests on a presumptively equal footing.[46] The Article 8(1) rights of each interested individual to respect for his or her family life must be considered in accordance with the requirements outlined in Article 8(2). The individuals' rights will subsequently be weighed against one another in a final balancing exercise, in which the child's

[40] Eekelaar (n 2).
[41] ibid 242.
[42] ibid 243.
[43] ibid 244.
[44] ibid 245.
[45] Choudhry and Fenwick (n 1) 455.
[46] ibid 479.

welfare is privileged but not automatically decisive. In this way the parallel analysis harmonises with the reasoning process and predicted outcomes on Eekelaar's approach.[47]

While both the modified least detrimental alternative analysis and the parallel analysis are preferable to the welfare principle, we recommend the latter approach over the former. The first axis of comparison is relative ease of application. While the modified least detrimental alternative analysis sets out to be a nuanced approach to the parent-child dynamics, the fact that it is multi-pronged and involves ambiguous concepts such as 'significant life goals' may pose insuperable difficulties when it comes to application to actual proceedings in all their complexity. The second axis of comparison is in terms of overall coherence in the law. Given that the parallel analysis, but not the modified least detrimental alternative analysis, is modelled on the approach under the ECHR, it better harmonizes with the overall law.

B. Procedural Safeguards in Court Proceedings

Apart from ensuring that parents with disabilities are not indirectly discriminated against under the welfare principle, it is also important to ensure that there are sufficient procedural safeguards for parents with disabilities in the relevant court proceedings. For example, in *RP v UK*[48] the ECtHR held that, having regard to the Article 6 right to a fair trial enshrined in the ECHR, a person must be appointed to represent the best interests of a parent with learning disabilities in court proceedings.

I. Checklist

Pursuant to section 7 of the CA 1989, a court considering any question with respect to a child under the Act may solicit independent evidence on matters relating to the welfare of that child and may ask an appointed social worker, a family court adviser, or such other person as is appropriate to prepare a welfare report. Further, before an application for a care order is made, a local authority has a prior duty under section 47 of the CA 1989 to investigate cases that give them reasonable cause to suspect that a child is suffering, or is likely to suffer significant harm. Alternatively, a local authority may have an obligation to investigate cases when directed by a court to do so under section 37 of the 1989 Act.

While there is broad policy guidance on how welfare assessments and investigations are to be conducted,[49] and due credit must be given to social workers and other appointed persons as regards their practical experience in carrying out these assessments and investigations, such guidance may be inadequate in eliminating discrimination against parents with disabilities. First, the judgment calls these personnel will have to make can be especially complex in cases involving disabled parent(s). For example, a social worker may not fully understand how a disability impairs a parent's ability to care for the child, or she or he may lack the ability to perceive that the parent can provide a different but still adequate level of care for the child.[50] Secondly, the risk of inadvertent discrimination against a disabled parent is especially palpable when these personnel become risk-averse following social outrage over the workings of the child protection system in the aftermath of tragic incidents such as the Climbié[51] and Baby P[52] cases.

---

[47] Choudhry and Fenwick (n 1) 472.
[48] [2012] ECHR 1796.
[49] Department of Health, *Framework for the Assessment of Children in Need and Their Families* (Her Majesty's Stationery Office 2000).
[50] For a success story, see Sharon Dale, 'Interview: That's my boy: how paralysed mother proved experts wrong' *The Yorkshire Post* (10 June 2011) <http://www.yorkshirepost.co.uk/news/interview-that-s-my-boy-how-paralysed-mother-proved-experts-wrong-1-3461043> accessed 15 March 2016.
[51] See Lord Laming, *The Victoria Climbié Inquiry: Report of an Inquiry by Lord Laming* (Cm 5730, 2003).
[52] Katherine Sellgren, 'Baby Peter was failed by all agencies' *BBC News* (26 October 2010) <http://www.bbc.co.uk/news/education-11621391> accessed 15 March 2016.

It is suggested that the welfare assessments and investigations will be more thorough, efficacious, and fair if there were a checklist to structure the assessment. Subject to further studies in this direction, such a checklist could be placed on a statutory basis so as to empower the courts to monitor the manner in which the assessments are conducted. The degree to which a social worker adheres to the items on the checklist may also be relevant to the degree of deference given by the court to his or her assessment, though the courts should refrain from taking a formalistic approach to the application of such a statutory checklist. The weightage given by the social worker to each of the elements on the checklist may similarly be subject to review.

II. Duty to Give Reasons

In addition, local authorities should have a legal duty to give reasons in making applications for care orders. As Lord Mustill noted in the seminal case of *R v Secretary of State for the Home Department, ex p Doody*, the lack of a general duty to give reasons for an administration decision need not detain us since 'it is equally beyond question that such a duty may in appropriate circumstances be implied'.[53] The imposition of a duty to give reasons would promote transparency in the child protection system and could potentially reveal subconscious discrimination against parents with disabilities on judicial review. In order not to impose too onerous an administrative burden on the local authorities, the reasons need be no more than a concise account of the way in which they have arrived at their decisions.

6. *Social Solutions*

The legal solutions proposed above would better combat direct and indirect discrimination against parents with disabilities if they were supplemented with social solutions.

A. Adequate Support Structures

The need for the State to render adequate social support to parents with disabilities in meeting their child-rearing responsibilities cannot be overstated. Article 23(2) of the CRPD provides that 'State Parties shall render appropriate assistance to persons with disabilities in the performance of their child-rearing responsibilities'. Further, the Supreme Court stated in *Re B (A Child) (Care Proceedings: Threshold Criteria)*[54] that, if needed, disabled parents should receive the support they need to be adequate parents. In the same vein, in *Kent County Council v A Mother*, Baker J acknowledged that 'people with learning disability may, in many cases, with assistance, be able to bring up children successfully'.[55] Financial support, in the form of Independent Living Fund Payments, would be a crucial source of support for parents with disabilities but there is definitely room for other forms of support. Apart from support from the State, it is arguable that the wider community[56] ought to contribute to the empowerment of parents with disabilities as well. These support structures can play a very significant role in helping parents with disabilities provide the same level of care to their children as their non-disabled counterparts.

B. Partnership

A fundamental tenet of the CA 1989 is that it seeks to promote a sense of partnership between parents and local authorities. Thus, section 17(1)(b) expressly states that it shall be the general duty of every local authority 'so far as is consistent with that duty, to promote the upbringing of such children by their families by providing a range and level of services appropriate to those children's needs'. The benefits of a sense of partnership are manifold in relation to parents with disabilities. For example, parents are likely to be apprehensive about social workers and their intermittent spot checks,

[53] [1994] 1 AC 531, 564.
[54] *Re B* (n 13).
[55] [2011] EWHC 402 (Fam) [132].
[56] CRPD, Art 23(5).

particularly where their disability is psychiatric in nature. A welfare assessment can be conducted in the least intrusive manner so that the parents are not unduly stressed. As Herring noted, the ultimate aim is that parents having difficulties in parenting will perceive the local authority as a reassuring source of support and assistance rather than as a body to be feared.[57]

7. *Conclusion*

This article has sought to address an important but often overlooked question: how should the law balance the best interests of children with the obligation of the State to eliminate discrimination against persons with disabilities? It has set out the current approach taken by the courts as embodied by the welfare principle, and has concurred in the conclusion that parents with disabilities should be expected to provide the minimum level of care to their children as their able-bodied counterparts. However, the article has also considered alternative ways to balancing the interests of the parent and the child that could render parents with disabilities less susceptible to losing care or custody of their children. Further, it has proposed other legal solutions to combating direct and indirect discrimination against parents with disabilities, including procedural safeguards, a statutory checklist, and the imposition of a legal duty on local authorities to give reasons for making an application for a care order. Last but not least, the article has emphasized the importance of social solutions in this regard, such as adequate support structures provided by the State and the wider community, as well as bolstering the partnership approach between local authorities and parents. It is the hope of the authors that more can be done to protect the interests of parents with disabilities in keeping their families together.

[57] Herring (n 3) 597.

A Thing Less Likely to Do Mischief: An Alternative Application of the Rule in *Rylands v Fletcher* to 'Ultra-Hazardous' Activities

Ryan Chan-Wei*

Introduction

Through the past century, the rule in *Rylands v Fletcher*[1] has been the subject of much debate, to the point where Francis Bohlen commented that yet another analysis of *Rylands v Fletcher* would merely be 'a thrashing out of old straw'.[2] Considering how crowded the field is, it might be difficult to see what value yet another article can add, let alone one by an undergraduate. However, I humbly suggest that there is still more to be said, especially with respect to how *Rylands v Fletcher* has been applied in the specific context of 'ultra-hazardous' activities. In particular, the English courts have not adopted a distinct approach[3] in applying the rule in *Rylands v Fletcher* to 'ultra-hazardous' activities despite valid concerns raised to the contrary,[4] instead preferring to subject both 'ultra-hazardous' and non-ultra-hazardous' activities to the same *Rylands v Fletcher* standard. This article will demonstrate that such a position can, and should, be rethought.

1. *Overview*

The existing debate concerning *Rylands v Fletcher* liability for 'ultra-hazardous' activities has largely been framed as a stark choice between two polarities, to the point where it was once described as 'a storm centre'.[5] On the one hand, some see *Rylands v Fletcher* as an opportunity to impose a blanket rule of strict liability for 'ultra-hazardous' activities, by embracing the enterprise liability approach that has significantly shaped American tort law, and leaving the non-'ultra-hazardous' *Rylands v Fletcher* cases to negligence.[6] On the other hand, some see the principle as obsolete, arguing instead that the rule in *Rylands v Fletcher* is an anachronism in light of the development of the fault principle and that it should be subsumed into the broader tort of negligence, even in the context of 'ultra-hazardous' activities, as the Australian courts have done.[7]

In contrast, this article will suggest that a middle ground can be found. An over-extension of the rule in *Rylands v Fletcher* (i.e. by imposing a general regime of strict liability for 'ultra-hazardous' activities) is just as undesirable as subsuming it entirely into the law of negligence, as both are sweeping responses to a category of cases that often have unique fact situations and instead require a

* St. Anne's College, Oxford. I am deeply grateful to Dr Imogen Goold and Benjamin Mak for their comments on an earlier draft. I would also like to express my heartfelt thanks to the editorial team of the Oxford University Undergraduate Law Journal for reviewing this article. Any and all errors remain my own.

[1] [1868] LR 3 HL 330 (HL).

[2] Francis Bohlen, 'The Rule in *Rylands v Fletcher*' (1911) 59 U Pa L Rev 298, 373 and 423.

[3] This reluctance is primarily underpinned by concerns over the 'uncertainties and practical difficulties' of pursuing a separate approach for 'ultra-hazardous' activities, as highlighted by Lord Goff in *Cambridge Water Co v Eastern Counties Leather Plc* [1994] 2 AC 264 (HL) 305. These concerns will be addressed in Section 3.

[4] For instance, see John Fleming, *The Law of Torts* (9th edn, Law Book Company 1998) 341. Arguing in favour of strict liability, he criticizes *Read v Lyons* [1947] AC 156 (HL) for prematurely stunting the 'development of a general theory' of *Rylands v Fletcher* liability for 'ultra-hazardous' activities.

[5] Cecil Wright, 'The Adequacy of the Law of Torts' (1961) CLJ 44, 52.

[6] *Cambridge Water Co* (n 3) 304.

[7] *Burnie Port Authority v General Jones Pte Ltd* [1994] HCA 13.

more calibrated approach. For instance, it will be shown that the Australian approach has the right normative instinct (i.e. to focus on fault),[8] but the wrong practical implementation (i.e. the evidential requirement of proving duty of care and breach of that duty to make out the tort of negligence may be too onerous a burden for the defendant to bear).

Instead, *Rylands v Fletcher* liability should be kept as it stands in English law, with the exception that the test for 'ultra-hazardous' activities should be separate, because such activities carry a uniquely high risk of harm. Unlike ordinary activities, when engaging in an 'ultra-hazardous' activity, even a small degree of fault can lead to the manifestation of an inordinate amount of harm; there is thus value in treating them as distinct, in order to better account for the disproportionate risks they pose.

This article will therefore propose an alternative framework for assessing *Rylands v Fletcher* liability in the context of 'ultra-hazardous' activities, which consists of 1) a working definition of an 'ultra-hazardous' activity and 2) a 'reduced fault' liability test. It may not be a flawless approach in practice, but it is arguably better for the courts to adopt an untested framework that attempts to offer a more nuanced and context-specific analysis of individual cases, than to apply a rigid, broad-brush approach which swings to doctrinal polarities and ignores the factual differences across the broad spectrum of *Rylands v Fletcher* cases. After all, in the former, we are approximately right; in the latter, we are precisely wrong.[9]

2. *Understanding the Rule in* Rylands v Fletcher

In *Rylands v Fletcher*, the factual scenario was fairly straightforward. The defendants constructed a reservoir in order to provide water for their mill, with the permission of the owner of the adjacent land. They employed independent contractors to do so and they themselves took no part in the construction. While the defendants were not guilty of any negligence in their selection of the site for the reservoir, it eventually transpired that there were some long abandoned mine workings underneath the site. In particular, there were several old shafts, which appeared to have been filled in. When excavating the bed of the reservoir, the contractors came upon these shafts, but did not make their existence known to the defendants. It was found that the contractors were negligent in proceeding with the construction of the reservoir despite their knowledge of the shafts.

The plaintiffs were lessees of a mine that lay under land close to the reservoir. In working their mine, they had broken into some of the aforementioned 'long abandoned mine workings'. After the contractors completed work on the reservoir, the defendants filled it with water. The shafts (which were filled in by soil) gave way under the pressure of the water, allowing the water to flow into the abandoned mine workings. The water then flowed into an opening the plaintiff had made when breaking into the abandoned mine workings, and from there the water finally flowed into the plaintiff's mine.

At trial, the main point of contention was whether the defendants could be held liable irrespective of proof of negligence on their own part, or on the part of the independent contractors. Although the court had some doubts about whether an isolated escape, as opposed to a continuous state of affairs, could found an action in nuisance, the defendants were ultimately held liable.

In delivering the opinion of the Court of Exchequer Chamber, Blackburn J laid down the

[8] The complex jurisprudential debate surrounding the role of corrective justice and strict liability in tort law is beyond the ambit of this article. For the sake of clarity, it will suffice to note that the former should be preferred in context of *Rylands v Fletcher* liability. For more details as to why the latter should be approached with caution, see Lord Goff's judgment in *Cambridge Water Co* (n 3) 305-306. Also see John Fleming, *The Law of Torts* (9th edn, Law Book Company 1998) 369.

[9] To paraphrase a quote from famed investor Warren Buffett, who once said, 'It is better to be approximately right than precisely wrong.'

broad principle (subsequently affirmed by the House of Lords on appeal), which has come to be known as the rule in *Rylands v Fletcher*, that: 'the person who for his own purpose brings on his lands and collects and keeps there anything likely to do mischief if it escapes, must keep it in at his peril, and if he does not do so, is *prima facie* answerable for all the damage which is the natural consequence of its escape.'[10] To this, two more requirements have been added as the law developed: the land must be used in a way that is non-natural[11] and there must be foreseeability of damage of the relevant type.[12]

It is worth noting that although *Rylands v Fletcher* liability is sometimes mentioned in the same breath as strict liability, it is not actually strict liability in the traditional sense, which holds the defendant liable *regardless of fault*. In English law, liability under the rule in *Rylands v Fletcher* is currently only 'strict' in the much looser sense that the defendant may be held liable notwithstanding that he has exercised all due care to prevent the escape from occurring. Ultimately, a fault element (of unreasonable risk-taking) still matters, and this will be explored further in Section 4.

3. *Defining an 'Ultra-Hazardous' Activity*

A. Context and Existing Definitions

Since the alternative framework proposed for assessing *Rylands v Fletcher* liability in the context of 'ultra-hazardous' activities will be formed by a two-stage test, it is important to first delineate the meaning of an 'ultra-hazardous' activity. This section will demonstrate that, despite weighty arguments to the contrary, it is in fact possible to obtain a satisfactory working definition.

The arguments against adopting a definition of 'ultra-hazardous' activities are well known, but they are ultimately unconvincing. For instance, academics such as Peter Cane have suggested any attempts at defining 'ultra-hazardous' will always be problematic because the term is inherently too subjective. He argues that 'ultra-hazardous' conveys the idea of an unacceptable risk, but that the 'level of risk we are prepared to accept in any particular activity depends on how highly we value that activity'.[13] In light of this, he hesitates to re-define the rule in *Rylands v Fletcher* in the context of 'ultra-hazardous' activities, arguing that 'this option is unattractive (…) because of the difficulty of defining the concept of an "ultra-hazardous" activity'.[14] A similar sentiment was echoed by Donal Nolan when he noted (quoting Lord MacMillan in *Read v Lyons*) that 'in a progressive world things which at one time were reckoned highly dangerous come to be regarded as reasonably safe',[15] and to attempt to frame a legal classification would therefore be 'impracticable'.[16] This is understandably so, because the law needs to account for evolving public attitudes on what constitutes a 'highly dangerous' activity, rather than rigidly operating a definition that is too inflexible to accommodate these changes.

In the same vein, the House of Lords in *Cambridge Water Co* endorsed a Law Commission report[17] which expressed serious misgivings about 'a general concept of "ultra-hazardous" activity' because of 'uncertainties and practical difficulties of its application'[18] that arise from its subjectivity, with Lord Goff indicating that judges should be 'reluctant to proceed down that path'[19] if the Law Commission was unwilling to do so. This was underpinned by the reasoning that if a general rule

[10] *Fletcher v Rylands* [1866] LR 1 Ex 265 (Exch Ch) 279.
[11] *Rylands* (n 1) 339.
[12] *Cambridge Water Co* (n 3) 301.
[13] Peter Cane, 'The Changing Fortunes of *Rylands v Fletcher*' (1994) 24 U W Austl L Rev 237, 237.
[14] ibid.
[15] Donal Nolan, 'The Distinctiveness of *Rylands v Fletcher*' (2005) 121 LQR 421, 448.
[16] *Read v Lyons* [1947] AC 156 (HL) 172.
[17] Report of the Law Commission on Civil Liability for Dangerous Things and Activities (Law Com No. 32, 1970).
[18] *Cambridge Water Co* (n 3) 305.
[19] ibid.

concerning 'ultra-hazardous' activities cannot be applied with precision, it is probably best not to extend the rule in *Rylands v Fletcher* in English law, and instead leave regulation to statute, which can lay down more 'precise criteria'.[20] On this view, a general rule runs the risk of riding roughshod over nuances in the differences between particular activities (e.g. changing societal perceptions), by imposing liability on a very broad (and potentially indeterminate) swathe of activities. In contrast, statutory regulation supposedly offers more precision because it imposes liability on a clearly defined set of activities that Parliament has deemed to be 'ultra-hazardous', after taking into account nuances (e.g. context-specific norms) that are uniquely relevant to determining whether an activity is 'dangerous'.

However, while the aforementioned arguments are not without merit, it is ultimately unwise to leave the governance of 'ultra-hazardous' activities purely to statutory regulation, and not attempt to define them at all.

Firstly, a list stipulating which activities are 'ultra-hazardous' cannot cover infinite possibilities, so it is very likely that there would be gaps in the law. At times, such a *numerus clausus* approach may be appropriate (e.g. when delineating rights that can be created in property); however, for the sake of efficiency, it would be preferable to adopt a working legal definition that can be applied consistently and flexibly to all fact situations, and the subsequent paragraphs will demonstrate that such an analytical framework can indeed be found.

Secondly, with respect, coming up with a suitable definition might not be quite as impractical as Cane and Nolan make it out to be. Admittedly, there is inherent subjectivity in attempting to determine what level of risk is 'acceptable', but this article will also go on to show that any uncertainty that might arise could be tempered by the subsequent application of an objective standard (such as that of a reasonable man in the shoes of the defendant) in determining whether the risk is acceptable in a particular context.

In contrast to the English perspective, American jurists have adopted a markedly different approach in attempting to come up with a clearer definition of 'ultra-hazardous', such as in Section 520 of the First Edition of the Restatement of Torts, where an activity is deemed to be 'ultra-hazardous' if it: a) necessarily involves a risk of serious harm to the person, land or chattels of others which cannot be eliminated by the exercise of utmost care and b) is not a matter of common usage.

What they have proposed is far from perfect, but it has at least convincingly demonstrated that it is indeed possible to have a working definition of an 'ultra-hazardous' activity. The concerns raised by English academics and judges about the inherent subjectivity of defining an 'ultra-hazardous' activity are valid, to the extent that the definition of 'acceptable danger' is constantly evolving. However, the American jurists rightfully differ in their belief that nuances such as changing societal perceptions can still be accounted for under an objective standard. On the whole, their proposed approach is a useful definitional guide, although there are certain aspects that would benefit from greater clarity or a more context-specific approach.

For instance, three fundamental aspects of this definition arguably call for substantial clarification. Firstly, what degree of risk is required? Secondly, to what extent does that risk have to be foreseeable? Thirdly, what is meant by 'common usage'?

The lack of clarity surrounding the notion of 'common usage' is particularly concerning, as it is both vague and incoherent. It is vague because it is not clear whether 'common' refers to the public generally or just a class of people; even if it referred to a class of people, it would still be necessarily to clarify what the minimum size of such a class may be, and how 'common' the activity has to be within the context of that class.

[20] *Cambridge Water Co* (n 3) 305.

However, a greater cause for concern is the incoherence of the notion that an 'ultra-hazardous' activity by definition has to be one 'not of common usage'. From a policy standpoint, it is hard to see why this should be, because there are clearly examples of activities in 'common usage' that one might instinctively regard as 'ultra-hazardous'. For instance, nuclear power plants have been recognized as 'ultra-hazardous' by the Nuclear Installations Act 1965, but it is plausible that, at one point, those living near Three Mile Island or Fukushima (or any city that depends on nuclear power to support its economy and energy needs) would have regarded nuclear power plants as being of 'common usage', in the sense that it was 'customarily carried on by many people in the community'.[21] Consider also the use of explosives to blast open a mine, which is likely to be regarded as an 'ultra-hazardous' activity. In an urban centre, this would clearly not be of 'common usage'; however, by contrast, in a mining town, it is likely that this would be considered 'common usage' and not an unacceptable risk, because many residents of that town are likely to be involved in the mining activities too (and probably accustomed to the occasional debris flying into their property from the explosions).[22]

Nevertheless, Section 520 of the First Edition of the Restatement of Torts can still serve as a useful definitional guide, even though it is somewhat unclear, because its emphasis on 'common usage' rightfully draws attention to the notion that the level of risk considered to be 'unacceptable' for the purposes of an 'ultra-hazardous' activity should depend on the context. Instead of imputing a consistent objective standard of 'common usage' to all fact situations, which is a rather broad-brush approach; a more context-specific approach would be beneficial, because the extent to which an activity is of 'common usage' can depend significantly on the norms of a particular locality.

B. Proposed Definition

The proposed definition builds on the foundations set out in the First and Second Editions of the Restatement of Torts, in particular by modifying the test from the Second Restatement. It should be noted that although the Second Restatement uses the term 'abnormally dangerous' activity instead of 'ultra-hazardous' activity, this is largely a matter of semantics. 'Abnormally dangerous' is meant to be a slightly more accurate descriptor of how such activities tend not to be of 'common usage', but for practical purposes it can be used inter-changeably with 'ultra-hazardous', as was done in the cases of *Langan v Valicopters*[23] and *Bennett v Larsen*.[24]

It is also important to note that this stage of determining whether an activity is 'ultra-hazardous' in the first place is a necessary but insufficient prerequisite for imposing liability; liability only attaches in the second stage, after applying the 'reduced fault' liability test (which will be discussed in the next section). If an activity does not fulfil the 'ultra-hazardous' test in the first place, then the liability test subsequently applied will be that of the normal fault test under the tort of negligence or nuisance, rather than the 'reduced fault' test.

The test proposed in this article is modified to account for more context-specific criteria and to offer more clarification on issues such as the risk of harm. My submission is that the defendant would be engaging in an 'ultra-hazardous' activity if a reasonable man in the defendant's shoes would conclude that:

i. The activity is inherently dangerous, in that it involves a high risk of severe harm that cannot

---

[21] Section 520(e) of the First Edition of the Restatement of Torts and Section 520(i) of the Second Edition of the Restatement of Torts.

[22] It is important to note that on occasion 'there will be a regime of statutory liability in force, which will displace the common law rule, and frequently exceed it', as observed by Jenny Steele in *Tort Law* (3rd edn, Oxford University Press 2014) 697. For instance, consider Section 209 of the Water Industry Act 1991. However, barring the presence of legislation, it is still necessary to craft a working definition of an 'ultra-hazardous' activity that is broad enough to cover the examples above, lest they fall through the cracks.

[23] *Langan v Valicopters Inc* 567 P 2d 218 (Wash 1977).

[24] *Bennett v Larsen Co* 348 N W 2d 540 (Wis 1984).

be eliminated despite the exercise of reasonable care;

ii. And the activity is not of 'common usage' (in the sense that it is neither customarily nor contextually appropriate to the locality in which it is carried out), such that the risks that accrue from it are deemed 'unacceptable'.

The first limb accounts for the higher risk of harm threshold inherent in 'ultra-hazardous' activities. Additionally, it highlights a key distinguishing factor between liability for 'ultra-hazardous' activities and ordinary negligence liability, which is that liability under the former can still be incurred notwithstanding the exercise of reasonable care. In contrast, it is unlikely that ordinary negligence liability would be found if the harm could not be eliminated despite the exercise of reasonable care, as it would not be 'fair, just and reasonable' (under the third limb of the *Caparo*[25] test) for the courts to implement a duty of care in that context.

The second limb addresses variations in context that are not accommodated by the Second Restatement's original definition, as discussed earlier, because the extent and type of risk that one might deem abnormal or 'unacceptable' ultimately depends on the context. Furthermore, it also addresses the possibility that the same activity can carry a different risk of harm, depending on the circumstances. For example, storing nuclear waste in an underground bunker in the middle of the desert carries a lower risk of harm than storing it in an underground bunker next to a densely populated city.

The test might be subject to criticism on the basis that its objectivity is merely a veil for the judiciary to apply a policy decision, as an objective standard tends to be inherently vague and how it is applied depends largely on who decides the case. Nevertheless, this has not prevented a workable objective test from being applied in numerous other circumstances (e.g. negligence and product liability), and thus should not pose a significant obstacle in this instance either.

4. *An Alternative Application of the* Rylands v Fletcher *Rule to 'Ultra-Hazardous' Activities*

A. Lord Bingham's *Obiter Dictum* in *Transco*[26]

There is value in treating 'ultra-hazardous' activities as distinct, because they carry a uniquely high risk of harm. Unlike ordinary activities, when engaging in an 'ultra-hazardous' activity, even a small degree of fault can lead to the manifestation of an inordinate amount of harm; and such activities should therefore be afforded a separate test in order to better account for the disproportionate risks they pose. Although the English courts have been hesitant to treat 'ultra-hazardous' activities as distinct, due to difficulties in defining 'ultra-hazardous', the preceding section has shown that it is possible to come up with a working legal definition. Having addressed that, this article will now explore how Lord Bingham's *obiter dictum* in *Transco* could hypothetically be applied, by extrapolating from his passing remarks in that case to create a legal framework for thinking about *Rylands v Fletcher* liability in the specific context of 'ultra-hazardous' activities.

In *obiter*,[27] Lord Bingham observed that there was a 'category of case, however small it may be' in which it 'seems just' to extend the rule in *Rylands v Fletcher* and impose liability 'even in the absence of fault'. He did not offer any concrete explanations, but instead merely offered a few cases as examples of what would constitute this category, such as *Cory Brothers*,[28] *Rainham Chemical Works*[29] and *Cambridge Water Co*, in which the defendants were all arguably engaged in some form of 'ultra-hazardous' activity. However, although Lord Bingham referred specifically to an 'absence of fault', a common thread running through each of these cases is in fact the presence of a fault element

[25] *Caparo Industries plc v Dickman* [1990] UKHL 2, [1990] 2 AC 605.
[26] *Transco Plc v Stockport MBC* [2003] UKHL 61, [2004] 1 AC 1 [6].
[27] ibid.
[28] *Attorney General v Cory Brothers & Co Ltd* [1921] 1 AC 521 (HL).
[29] *Rainham Chemical Works v Belvedere Fish Guano* [1921] 2 AC 465 (HL).

of unreasonable risk-taking. However, the degree of fault in these cases is too insignificant to be captured by traditional, currently legally recognized standards of fault (such as negligence).

For instance, Lord Bingham suggests that liability should have been imposed in *Cory Brothers*, even if the claimants had failed to prove negligence, since 'memories of the tragedy at Aberfan are still green'.[30] The facts of *Cory Brothers* are similar to that of the Aberfan tragedy, where a waste tip slid down a mountainside into a village and killed many people, as the defendants placed up to half a million tons of mineral waste on a steep hillside, some of which eventually slipped off and damaged nearby properties. Lord Bingham appears to be suggesting that liability in the absence of negligent fault might be appropriate because the defendant was aware of the potential for severe harm to occur, and is therefore at fault for proceeding with the activity without making any effort to take precautions to mitigate the risk of harm materializing. This echoes the view of Viscount Haldane in *Cory Brothers*, who argued that instead of proving negligence (as the claimants did), it would suffice merely to 'show that the hillside was steep, and that to pile rubbish on it in a large heap was to put a dangerous structure there'.[31]

Similarly, in *Rainham Chemical Works*, the defendants were involved in manufacturing picric acid (an explosive chemical) for the British government to make munitions during the First World War. However, they were not experts in chemistry and did not realize that that the picric acid was in fact highly explosive. Although there was no proof of negligence in their production process, they were nonetheless held liable for damage to neighbouring property when a fire occurred in their factory and the picric acid led to an explosion of 'terrific violence'.[32] A fault element is less obvious than in *Cory Brothers*, but was arguably still present: the chemical clearly had the potential to cause serious harm (the defendants knew that they were being manufactured as part of a war effort to make explosive munitions), and the defendants could be faulted for not attempting to determine the exact extent to which their chemicals were explosive. A reasonable man would arguably have made an effort to identify the specific explosive potential of the chemical, before proceeding with their production, in order to be better placed to take precautions to prevent the risk of harm from materializing.

Finally, Lord Bingham held that an extension of the rule in *Rylands v Fletcher* would be justified in *Cambridge Water Co* if the damage had been foreseeable by a reasonable man, and it is clear a fault element would have been present on the part of the defendant tannery in such a case, as a reasonable man would have taken more steps to prevent their industrial solvents from percolating down to the water table – after all, it was established in the case that the solvents were 'bound to cause mischief' if they escaped.[33]

It is important to note that although Lord Bingham might appear to be proposing the imposition of a strict liability regime (albeit limited to the abovementioned fact situations), since he mentions the imposition of liability 'even in the absence of fault',[34] this is unlikely to be the case. It would be inordinately difficult to justify the imposition of a *limited* strict liability regime,[35] as any arguments used to establish this (e.g. economic efficiency) are likely to apply with equal force to the

---

[30] *Transco Plc* (n 26) [6].
[31] *Attorney General* (n 28) 536.
[32] *Rainham Chemical Works* (n 29) 471.
[33] *Cambridge Water Co* (n 3) 288-289.
[34] *Transco Plc* (n 26) [6].
[35] Even if a case could successfully be made for the imposition of a limited strict liability regime (e.g. 'only in the absence of fault'), it would then be very difficult to justify circumscribing it only to that particular context, because of the all-encompassing nature of strict liability, which imposes liability *regardless* of whether the defendant was at fault or not. Therefore, it is likely that any arguments for a *limited* strict liability regime would apply with equal force to a *general* strict liability regime, because the former is not easily distinguishable from the latter.

imposition of a *general* strict liability regime,[36] and Lord Bingham is clearly against the latter, believing it is properly a matter for Parliament and not the courts to decide.[37]

B. Proposition: 'Reduced Fault' Liability Test

So far, it is clear from Lord Bingham's *obiter dictum* that a degree of fault may well be present in *Rylands v Fletcher* cases that involve 'ultra-hazardous' activities. However, a new (reduced) standard of fault should be adopted when assessing the liability of defendants, as the degree of fault in those cases is often too insignificant to be captured by traditional, currently legally recognized standards of fault.

Thus, extrapolating from the underlying thread of a fault element identified in the three cases mentioned earlier, liability could be imposed with respect to an objective test that measures fault with respect to foreseeability of risk of harm, with the fault threshold fulfilled if a 'reasonable person' in the defendant's shoes would have foreseen a high risk of harm materializing (e.g. personal injury, property damage etc.) and acted differently. There may not be any negligent breach of a duty of care, but *Rylands v Fletcher* liability should still be incurred on the basis that the defendant did not adequately address the inherent dangers present in an 'ultra-hazardous' activity.

The lowered fault requirement applied in this test does not have a direct equivalent in tort law, but a parallel could be drawn to the *Caldwell* test,[38] which was formerly the test for recklessness in criminal law, especially if the second limb of that test was modified to make it objective as well. In essence, the modified variant of the *Caldwell* test applied in this context would impose liability on a defendant for all damage which is not too remote a consequence of his actions (following the *Wagon Mound*[39] principles, as applied in *Hughes v Lord Advocate*[40]) if:

i. The defendant does an act, which he should have reasonably foreseen as leading to a high risk of harm materializing,
ii. And a reasonable man in the defendant's shoes would not have taken such a risk.

The ostensible harshness of this objective test towards the defendant is balanced out by the fact that it can only be applied once the activity in question has fulfilled the 'ultra-hazardous' test, whose context-specific analysis sets a high bar for arguing that an activity is 'ultra-hazardous' in the first place. It is a difficult threshold to fulfil, because it must be shown that the risks were 'unacceptable' in a very specific context (and not just generally, which would have been easier to establish), and this acts as a control mechanism to ensure that only defendants engaging in activities with a markedly unacceptable level of risk are deservedly subject to the more stringent 'reduced fault' test.

At any rate, it is important for this test (and the 'ultra-hazardous' test as well) to be an objective one assessed with respect to the standard of the 'reasonable man' as opposed to a subjective test; this objectivity acts as a control mechanism[41] that judges can use to lay down extra-legal standards without which the law 'cannot do its work in a sufficiently sensitive way'.[42] In the case of *Rylands v Fletcher* liability for 'ultra-hazardous' activities, such 'sensitivity' is necessary in order for judges to account for

---

[36] Regrettably, this article does not have the scope to delve further into the intricacies of strict liability; it will suffice to note that there are good reasons to be wary about the imposition of a strict liability regime, because, *inter alia*, it runs the risk of riding roughshod over key tort principles such as corrective justice. For further reading, consider 'Tony Honoré, *Responsibility And Fault* (Hart Publishing 1999)' and the corresponding Festschrift 'Peter Cane and John Gardner (eds), *Relating To Responsibility: Essays for Tony Honoré on His Eightieth Birthday* (Hart Publishing 2001)'.
[37] *Transco Plc* (n 26) [6].
[38] *R v Caldwell* [1982] AC 341 (HL).
[39] *Overseas Tankship Ltd v Morts Dock and Engineering Co Ltd* [1961] UKPC 2.
[40] *Hughes v Lord Advocate* [1963] UKHL 8, [1963] AC 837.
[41] John Gardner, 'The Many Faces of the Reasonable Person' (2015) 131 LQR 563.
[42] ibid, 4.

the varied impacts that differing contexts can have on the extent to which an activity is 'ultra-hazardous' and what a reasonable response to the risks present would be. An example would be that a defendant storing chemical waste is likely to be considered an 'acceptable risk' in an industrial town, but not in an urban centre, and therefore the level of precautions a defendant would be reasonably expected to undertake will be higher in the urban centre than in the industrial town. Unlike the Pearson Commission's suggestion[43] to arbitrarily draw a line in the sand and impose strict liability on certain activities by statutory instrument, this 'reduced fault' test offers a more calibrated response and does not ride roughshod over the impact that the context can have.

Applying the proposed test to the cases mentioned above, both limbs would have been fulfilled in *Cory Brothers*, as 1) it is clear on the facts that the defendant should have reasonably foreseen (and in fact did foresee) a risk of harm materializing from the unsecured mineral waste slipping off the heap and 2) a reasonable man in the shoes of the defendant would not have taken such a risk, as Viscount Haldane clearly states that 'it was *negligent* to put (the waste) there without taking adequate precautions [emphasis added]'. [44] Similarly, both limbs would have been fulfilled in *Rainham Chemical Works*, as 1) the defendants should have reasonably foreseen a risk of harm materializing from manufacturing an explosive chemical and 2) knowing of the risk of explosion, a reasonable man would not have proceeded with production without first determining the extent to which the chemical could pose a hazard, in order provide for adequate protective measures. It is not necessary for the defendant to have knowledge of all the circumstances before liability can be found, as long as the defendant is sufficiently aware of surrounding factors that should reasonably give rise to a cause for concern.

Ultimately, this alternative approach attempts to strike a balance between two extremes – the American approach, which has extended the rule in *Rylands v Fletcher* to impose strict liability for all 'ultra-hazardous' activities (though they now use the slightly different terminology of 'inherently dangerous activity'), and the Australian approach, that has subsumed the rule in *Rylands v Fletcher* into the tort of negligence.

On the one hand, the American approach imposes a broad swathe of liability, and no cases 'fall through the cracks', but the imposition of a general regime of strict liability comes at an inordinately significant cost in the context of *Rylands v Fletcher* liability (which this article unfortunately does not have the scope to discuss).[45] *Inter alia*, a strict liability regime is likely to disrespect our status as autonomous agents with inherent dignity because it may hold us liable for things we have not voluntarily chosen. By contrast, a fault-based approach respects our autonomy by ensuring that we are only liable for the things we have chosen by our own volition to do, and should be preferred in the context of *Rylands v Fletcher* liability for 'ultra-hazardous' activities.

On the other hand, the Australian approach avoids the drawbacks of a general strict liability regime, but requiring claimants bring a negligence claim instead sets the bar too high, making it too easy for a defendant to escape liability; the requirement to establish a duty of care and then breach of that duty may be an 'insurmountable evidentiary burden'.[46] For instance, as discussed earlier, in the three cases mentioned by Lord Bingham in *Transco*, the defendants were 'at fault' in some way, but not always to the point that they would have been liable in negligence.[47] Although the High Court of

---

[43] *Report of the Royal Commission on Civil Liability and Compensation for Personal Injury* (Cmd 7054, 1978) (The Pearson Commission).

[44] *Attorney General* (n 28) 536.

[45] As mentioned previously (n 8), for more on the dangers of a general strict liability regime, see Lord Goff's judgment in *Cambridge Water Co* (n 3), 305-306.

[46] John Murphy, 'The Merits of *Rylands v Fletcher*' (2004) 24 OJLS 643, 660.

[47] This was also the case in *Ilford Urban District Council v Beal and Judd* [1925] 1 KB 671 (KB), in which the source of the harm was not discoverable by reasonable care, hence there could not have been an actionable negligent breach of duty.

Australia was right to conclude in *Burnie Port Authority* that a duty of care analysis is appropriate since *Rylands v Fletcher* cases often reflect a relationship where the claimant is in a special position of dependence or vulnerability, it is precisely because of this vulnerability that it might not be appropriate for a *Rylands v Fletcher* case to be brought under a negligence action. After all, it is eminently conceivable that the disparity between the plaintiff and defendant could be so large (e.g. the defendant might be a large industrial enterprise) as to be 'highly improbable that a (relatively impecunious) claimant would be able to establish the necessary lack of reasonable care for the purposes of a negligence action'.[48] Essentially, the Australian approach has the right normative instinct (to focus on fault), but the wrong practical implementation (the evidential requirement of proving duty of care and breach of that duty is too onerous a burden for the defendant to bear).

In contrast, this alternative approach of applying a 'reduced fault' test attempts to tread the middle ground and offers a more calibrated approach. It lowers the bar from the Australian approach, such that claims can still be brought as long as some fault element can be established. At the same time, this does not impose a regime of strict liability as broad as the American approach, which may prevent cases from slipping through the cracks, but also brings with it a host of undesirable attendant consequences.

By way of illustration, consider two defendants – one storing hazardous chemicals, and another transporting nuclear waste. The former is aware that his chemicals are somewhat hazardous, but does not make an effort to determine exactly how hazardous they are, beyond placing them in ordinary storage barrels conventionally used for storing hazardous chemicals. However, the chemicals are more hazardous than he expects, eventually corroding the barrels and leaking out, causing severe property damage. In contrast, the latter takes precautions that reduce the level of risk to miniscule proportions, such as transporting the nuclear waste in an armoured train, using several layers of durable protective casts.[49] However, the waste leaks out when the train is derailed due to a fault in the tracks, and causes equally severe property damage.

According to the American approach, which imposes a general regime of strict liability for all 'ultra-hazardous' activities, it is likely that both defendants will be found liable. On the other end of the spectrum, according to the Australian approach, it is plausible that neither defendant will be found liable. In the latter case, there was no negligence on the part of the defendant; in the former, it may not be possible to prove negligence, as the defendant's fault may not have been sufficient to establish a breach of duty (as in *Rainham Chemical Works*).

However, under the proposed test, a more calibrated response can be found, such that the defendants' fault is taken into account and they are not treated with the same broad-brush approach despite their different factual situations. Applying this test to the first defendant, he would have been found liable, as it is 1) it is plausible that he should have reasonably foreseen a risk of harm materializing from storing hazardous chemicals and 2) knowing of the risk of harm, a reasonable man would have first determined the extent to which the chemicals could pose a hazard, before proceeding with storage, in order to ensure that adequate protective measures were taken to prevent their escape. In contrast, applying this test to the second defendant, he would not have been found liable. Even though 1) it is reasonably foreseeable that transporting nuclear waste might lead to a high risk of harm materializing, 2) a reasonable man in the defendant's shoes would have taken such a risk anyway, because of the extensive precautions that had been taken and the fact that he is unlikely to have had any way of knowing about a fault in the tracks.

Finally, a corollary of the 'reasonable man' element in the second limb of the test is also that a defendant will not liable for any damage that it would have been unreasonable to expect him to know

[48] John Murphy (n 46), 660.

[49] An example borrowed from Ken Kress, 'The Seriousness of Harm Thesis for Abnormally Dangerous Activities' in David G. Owen (ed), *Philosophical Foundations of Tort Law* (Oxford University Press 1995) 281.

about (e.g. damage arising due to unforeseen occurrences). By way of yet another example, consider a defendant engaging in an 'ultra-hazardous' activity such as using dynamite to blast a tunnel through a mountain (this carries with it the risk of landslides). As long as he has taken adequate precautions such as putting up warning signs well in advance and scanning the immediate vicinity to make sure there are no civilians present, he would not be liable if some hikers disregarded his warnings, set up a camp near the mountain at the last minute and consequently were injured or killed by the landslide resulting from the explosion. After all, it is plausible that a reasonable man in the defendant's shoes would have gone ahead with the activity anyway, as all necessary precautions were taken and there was no way to know about the presence of the hikers.

5. *Conclusion*

This article has proposed an alternative framework that analyzes *Rylands v Fletcher* liability for 'ultra-hazardous' activities in a way that seeks to tread the middle ground between the doctrinal extremes of the American approach and the Australian approach. Although untested and probably far from perfect, it is hoped that this framework is nevertheless a useful contribution to the existing debate on *Rylands v Fletcher*, because it endeavours to apply an 'approximately right' calibrated approach to a problem that has thus far been vulnerable to the 'precisely wrong' broad-brush approaches which swing to the doctrinal extremes of either imposing a general regime of strict liability, or subsuming the rule in *Rylands v Fletcher* entirely into the law of negligence.

The future after *Durant*: is backwards tracing the way forward?

Alexandra Clarke*

Introduction

In the seminal case of *Foskett v McKeown and Others*[1] (‘***Foskett***’) Lord Millet clearly explained that tracing is ‘neither a claim nor remedy. It is merely the process by which a claimant demonstrates what has happened to his property, identifies its proceeds and the persons who have handled or received them, and justifies his claim that the proceeds can properly be regarded as representing his property’. The rules governing the tracing process are more generous to claimants in equity than at law (for example, allowing tracing through a mixed fund). However, the fundamental principles are the same in either case: tracing is simply the process by which trust assets or their exchange products are identified following transactions.

It is for this reason that the possibility of backwards tracing has traditionally been rejected by the English courts. Also sometimes referred to as ‘tracing into the payment of a debt’, backwards tracing would allow the process of tracing to occur in the situation where an asset has been acquired in exchange for the creation of a debt and, subsequently, misappropriated trust assets are used to discharge that debt. Hence, it differs from traditional tracing which operates only where there is a substitute asset which has been legally acquired using the trust assets; in backwards tracing cases, the asset being traced has already been acquired *before* the trust assets are used. In the case of *Bishopsgate Investment v Homan*,[2] Leggatt LJ clearly sets out that ‘there can ordinarily be no tracing into an asset which is already in the hands of the defaulting trustee when misappropriation occurs’ because that asset ‘existed and so had been acquired before the money was received and therefore without its aid’. Although this statement was *obiter* and made without express agreement from the other members of the Court of Appeal, it reflects the typically unenthusiastic response of the English courts to the possibility of such backwards tracing.

However, since *Foskett* definitively clarified the identificatory purpose of tracing, the status of the doctrine of backwards tracing in English law has been unsettled. Despite receiving judicial endorsement in *obiter* in several cases,[3] it is yet to be formally recognized as permissible in English law. The main reason for this uncertainty is that there has not yet been an English case directly addressing this issue.

However, the recent Privy Council decision *The Federal Republic of Brazil v Durant International Corporation*[4] (‘***Durant International***’), on appeal from Jersey, did deal squarely with the subject of backwards tracing. The case involved a series of payments, which the Royal Court of Jersey found represented bribes, made through three different bank accounts. The dispute concerned the quantum of liability of Durant and Kildare as constructive trustees of the bribe money; the appellant companies appealed on the basis that their liability should be USD$7.7m, rather than USD$10.5m.

* Graduate of Selwyn College, Cambridge in July 2016; now studying the Bachelor of Civil Law at Magdalen College, Oxford. I am grateful to all who contributed to improving this piece. All mistakes are my own.

[1] [2000] UKHL 29, [2001] 1 AC 102.

[2] [1994] EWCA Civ 33, [1995] Ch 211, 221.

[3] See comments of Sir Richard Scott in Court of Appeal, *Foskett v McKeown* [1998] Ch 265, 284 and Lady Justice Arden in Court of Appeal, *Relfo Limited v Varsani* [2014] EWCA Civ 360, [63].

[4] [2015] UKPC 35.

Their argument relied on both the 'no backwards tracing' doctrine and the 'lowest intermediate balance' doctrine. It is the former submission which is pertinent for present purposes, although the two doctrines are inter-related.

The specific facts of the case are crucial. The full sum of the bribe proceeds, USD$10.5m, was paid in instalments into the first account, i.e. the 'Chanani account'. Before the USD$10.5m had been fully paid into the Chanani account, six payments totalling USD$13.1m were transferred from the Chanani account into a second account, i.e. the 'Durant account'. Lastly, there were four payments of USD$13.5m from the Durant account to the third and final account, i.e. the 'Kildare account'.

The respondents argued that the full USD$10.5m sum could be traced through to the Kildare account. However, the appellants pointed out that three of the payments into the Chanani account were made only after the sixth payment had already been made to the Durant account. Thus, there could be no tracing of those three payments into the Durant account because the chronology meant that they could not be regarded as substitute assets for the original bribe payments into the first Chanani account which had not yet been made. Accordingly, only the USD$7.7m which had been transferred in the correct order should be traceable.

Despite explicitly acknowledging that 'conceptually the appellants' argument is coherent and it is supported by a good deal of authority',[5] this reasoning was not accepted and the appeal was accordingly dismissed. Contrary to traditional English authority, Lord Toulson, on behalf of the Privy Council, 'reject[ed] the argument that there can never be backward tracing' and proposed a test based on 'coordination between the depletion of the trust fund and the acquisition of the asset which is the subject of the tracing claim'[6] to determine when it should be permitted.

The first part of this essay argues that *Durant International* was rightly decided to the extent that the law of tracing should be expanded to allow backwards tracing, albeit only in some, well-defined situations. This is because:

a. This position is consistent with the equitable principles of tracing, as described by the majority of the House of Lords in *Foskett*;

b. It is desirable from a policy perspective because:

    i. It is necessary to combat complicated financial fraud;
    ii. It would better serve justice to trust beneficiaries; and
    iii. It would not unduly prejudice unsecured creditors.

c. It is accepted that the current rigid, chronological approach to tracing provides legal certainty. However, it does not do justice to the parties and results in some arbitrary distinctions. Therefore, on balance, a more nuanced, purposive solution, which allows for backwards tracing in some circumstances, should be preferred.

However, the case of *Durant International* does not satisfactorily or fully illuminate the law on backwards tracing because the test propounded by Lord Toulson is too vague and imbues judges with too great a degree of discretion. The test requires structure and its boundaries more precise clarification. On the other hand, on one interpretation, the test may limit too greatly the circumstances in which backwards tracing can operate.

In light of this conclusion, the second part of this essay considers a range of different, more specific tests to determine when the process of tracing should be allowed to operate backwards. The conclusion of this discussion is that a test founded on the intention of the party acquiring the relevant

[5] *Durant International* (n 4) [18].
[6] ibid [40].

asset, as suggested by Sir Richard Scott in the Court of Appeal in *Foskett v McKeown*[7] is the preferable approach because:

a. Intention provides a solid and ascertainable link between the relevant transactions;
b. The courts are experienced at applying tests of intention;
c. Such a test is structured and so does not rely on unbounded judicial discretion; and
d. It is consistent with the flexible, commonsense approach applied to tracing in *Foskett.*

1. *Should backwards tracing be recognized?*

A. As a matter of principle: substance over form

In order to maintain coherence and consistency in this area of the law, any legitimate expansion of the current scope of the tracing process, including the recognition of backwards tracing as valid, should be consistent with the accepted principles of tracing. The most recent and authoritative elucidation of these principles is provided by the House of Lords in the seminal case of *Foskett*, the facts of which are well-known and so do not bear repeating here.

Lord Millet, with whom the majority agreed, makes it very clear that what is traced 'is not the physical asset itself but the value inherent in it'.[8] Furthermore, whether tracing can or cannot occur is a question of 'attribution not causation'.[9] Thus, by favouring a flexible approach over a strict, chronological approach and stressing transactional links, Lord Millet leaves scope for backwards tracing. Professor Graham Virgo is a proponent of this view: he points out that 'if tracing is not concerned with causation as such, in the sense that but-for the receipt of the claimant's property, the substitute asset would not have been obtained, but is concerned with attribution of value, it is surely possible to attribute value from the original asset to the substitute asset if the claimant's money has been used to discharge a debt incurred in respect of the substitute asset'.[10]

This discussion of debt discharge can be exemplified by the *Durant* case: not all of the money transferred from the Chanani account into the Durant account could be, strictly speaking, regarded as the bribe proceeds because not all of the bribe proceeds had yet been transferred into the Chanani account and it is trite law that *nemo dat quod non habet*, i.e. nobody can give what they do not have. Thus, when the rest of the bribe money did subsequently come into the Chanani account, rather than being transferred to the Durant account, it simply filled the gap left by the payment which had already been made; it effectively paid off a debt.

Virgo's argument is that, by abandoning causation as the justifying concept, it is not necessarily fatal to the process of tracing that the payments in *Durant* were made in the wrong order or, as another example, title to an asset passed before trust money was used to discharge the debt incurred.

In the opposite camp is Professor Matthew Conaglen.[11] He initially relies on the idea that tracing is a process whereby assets are identified and that the discharge of a debt 'is not the same thing as the acquisition of an asset'.[12] Therefore, it does not make sense to talk about 'tracing into a debt' for the purposes of backwards tracing.

However, this misses the point of the attribution model, which tracks the value of the asset. In the case of a discharge of a debt that has been incurred in the acquisition of an asset, this value is the amount of the debt, which can then be traced into the asset (the purchase of which incurred that debt)

[7] [1998] Ch 265, 284.
[8] *Foskett* (n 1) 128.
[9] ibid 138.
[10] Graham Virgo, *The Principles of Equity & Trusts* (1st edn, OUP 2012) 651.
[11] Matthew Conaglen, 'Difficulties with Tracing Backwards' (2011) 127 LQR 432.
[12] ibid 447.

as if it had been the purchase price. Conaglen does not accept this rationalization because he does not consider that there is 'a substitution of one asset for another asset (…) where one seeks to trace money (or other valuable property) into the repayment of a debt'.[13] However, this is because he has taken too narrow a view of the concept of substitution. In these backwards tracing cases, there is indeed a form of indirect substitution because when the debt is incurred to acquire title to the asset and the trust money is then used to discharge that debt, the end situation is the same as if that money had been used to acquire the asset in the first place. In other words, as Professor Lionel Smith puts it, the payment to discharge the debt is essentially 'delayed payment'[14] for the asset. This is what is often referred to as a 'substance over form' approach.

In fact, later on in his article, Conaglen appears to accept the possibility of such indirect substitution, saying that 'when the trustee uses trust funds to pay the debt, the beneficiaries' money has been used to pay the price of the asset' although 'only in a loose sense' because 'in return for the asset, the vendor accepted the debtor's obligation to pay the debt',[15] not the beneficiaries' trust money. This concession suggests that Conaglen's position is not in reality very far removed from that expressed by Smith, Virgo, and this essay.

In *Durant International* itself, the Privy Council seems to accept that the appellants' argument that 'a property interest cannot turn into (or provide a substitute for) something which the holder already has; the later acquisition cannot be the source of the earlier'[16] is coherent and supported by authority. Nevertheless, the court ultimately rejects it, and embraces the possibility of backwards tracing, which implies that they do recognize some form of substitution, even where title has already passed. However, there is no in-depth discussion of this conceptual point, and Lord Toulson seems satisfied to treat Conaglen's acceptance 'that there is nothing conceptually impossible'[17] about backwards tracing as conclusive of the matter. The case therefore does not provide much insight on the present question of whether backwards tracing fits within the general conceptual framework of tracing.

Conaglen concludes that the adoption of backwards tracing 'by the courts would mark a significant departure from, or addition to, the traditional principles of the law of tracing'.[18] Whereas I agree that it was a significant step for the Privy Council to recognize backwards tracing (if only because there was no real prior authority for it), it was not a dramatic departure from the principles of tracing, as presented in *Foskett*, because, as demonstrated above, backwards tracing fits quite neatly within the flexible, attribution model.

On a slightly different note, Professor Andrew Burrows,[19] in a passage quoted by the Privy Council in *Durant International*, argued that '"backwards tracing" must be accepted if one is to explain tracing into and through "in credit" bank accounts', in light of the common banking practice of crediting accounts before the relevant transferred funds are actually received. In other words, he explains 'the debt owed by the bank to the customer, which is treated as a substitute for the funds, exists in advance of the funds being received'. Thus, English law should recognize backwards tracing, not merely because it is conceptually sound or normatively desirable, but because it already regularly occurs. It is perhaps too wide a proposition to say that this provides a convincing basis on which to accept backwards tracing in all, or even a wide range of, circumstances; after all, the particular fact pattern, relating to online banking, is very specific and depends to a certain degree on the technology involved. However, Burrows' bank account example does serve to illustrate that the courts are willing

[13] Conaglen (n 11) 448.
[14] Lionel Smith, *Tracing into the Payment of a Debt* (1995) 54 CLJ 290, 292.
[15] Conaglen (n 11) 450.
[16] *Durant International* (n 4) [17].
[17] Congalen (n 11) 455.
[18] Conaglen (n 11) 449.
[19] Andrew Burrows, *The Law of Restitution* (3rd edn, OUP 2011) 142.

to depart from a strictly chronological approach where to do otherwise would run contrary to reality and produce absurd results. This further backs up my conclusion that recognizing backwards tracing is not a giant leap for the courts to take, but a small sensible step in the right direction.

B. As a matter of policy: fighting fraudsters and striking a balance

It has been demonstrated that English law on tracing would not have to significantly readjust in order to accommodate backwards tracing. However, this alone is not sufficient reason to recognize backwards tracing; as Conaglen rightly notes,[20] it must also be beneficial as a matter of policy. The two most pressing policy arguments relate to the fight against large scale financial crimes and ensuring a fair balance between the position of trust beneficiaries and the unsecured creditors of the trustee(s).

I. Combating financial crimes

In *Durant International*, the decisive argument in favour of permitting backwards tracing was that it is necessary in order to prevent criminals benefitting from complex financial transactions designed for the specific purpose of circumventing the law on tracing. As the Royal Court of Jersey pointed out, 'otherwise any sophisticated fraudster would be able to defeat an otherwise effective tracing claim simply by manipulating the sequence in which credits and debits were made to his account'.[21] Accordingly, a strict, chronological approach to tracing is highly undesirable, especially in the context of modern day commerce, which allows for not only much more sophisticated and complex transactions, but also much more complex and hard-to-detect crime. This was highlighted by Lord Toulson in *Durant International*: 'the development of increasingly sophisticated and elaborate methods of money laundering, often involving a web of credits and debits between intermediaries, makes it particularly important that a court should not allow a camouflage of interconnected transactions to obscure its vision of their true overall purpose and effect'.[22] This view is clearly right and perfectly consistent with the substance over form approach taken to tracing in *Foskett*.

However, this objective is only relevant in cases involving complicated financial transactions, as were present on the facts of *Durant International*. It does not apply to the more basic examples of backwards tracing, such as where a debt is incurred by the purchase of a car, which is then paid off with trust assets, because this could widen the scope of tracing too far and result in tracing even where the link between the two transactions is very weak. Thus, although strong in this particular context, this argument only supports backwards tracing in the limited situations involving the risk of financial crime.

II. Fair balance between beneficiaries and creditors

A more far-reaching concern relates to the impact of recognizing backwards tracing on the parties involved. Conaglen[23] particularly emphasizes that it would be to the detriment of unsecured creditors in the event of insolvency because it would effectively reduce the pool of the trust assets available to them (as beneficiaries would be able to carry out more far-reaching claims). This is undoubtedly true. However, it is not a very strong objection to recognizing backwards tracing; there are several possible counter-arguments

Firstly, not all situations in which backwards tracing could conceivably apply involve unsecured creditors. Often, as in *Durant International* itself, there is no risk of insolvency and therefore it is difficult to use the detriment argument to justify a bar on backwards tracing in those cases. However, Conaglen is right to point out that 'one of the primary reasons for trust beneficiaries wanting to be

[20] Conaglen (n 11) 446: 'a question of legal policy rather than principle'.
[21] *Durant International* (n 4) [13].
[22] ibid [38].
[23] Conaglen (n 11) 445.

able to trace backwards is in order to improve their recovery rate in the event of the trustee being made bankrupt'[24] because that is when a secured proprietary claim becomes most useful.

A second, more convincing argument, is that the majority of unsecured creditors are voluntarily unsecured; they may, for example, have turned down the opportunity to obtain security in exchange for a better price. As Smith explains 'such creditors could have taken security interests, which most commentators agree would give them priority over any tracing claim. If they did not, then they were exposed to the risk of tracing claims, like all unsecured creditors'.[25] This principle applies equally in the rare situations where the unsecured creditor gave the loan between the trustee's acquisition of the asset and the use of trust assets to pay off the debt used to acquire it (i.e. when it was impossible for the creditor to know that a proprietary right, able to trump their personal right, would later arise in favour of the beneficiaries). Moreover, 'there is no reliable way to determine whether an asset is subject to a tracing claim'[26] anyway, regardless of whether the process is operating backwards or forwards. Therefore, the creditor is unlikely to know about the beneficiaries' proprietary interest in any circumstances.

This concern for unsecured creditors, however, was taken seriously by the Royal Court of Jersey in *Durant International*. The Privy Council reported them as holding that 'at least where the account remained in credit during the relevant period, so there was no question of possible insolvency and prejudice to unsecured creditors, and where there was no suggestion of an intervening bona fide purchaser for value, the question should be whether there was sufficient evidence to establish a clear link between credits and debits to an account'.[27] Thus, consideration of the position of other parties was incorporated into their proposed 'clear link' test for backwards tracing.

Similarly, the Privy Council in *Durant International*, acknowledged that 'the courts should be very cautious before expanding equitable proprietary remedies in a way which may have an adverse effect on other innocent parties' and, more specifically, that 'if a trustee on the verge of bankruptcy uses trust funds to pay off an unsecured creditor to whom he is personally indebted, in the absence of special circumstances it is hard to see why the beneficiaries' claim should take precedence over those of the general body of unsecured creditors'.[28] Lord Toulson nevertheless decided that backwards tracing should, in principle, be permissible. This conclusion is difficult to reconcile with his concern for the unsecured creditors of trustees, making his position appear untenable. It would, for example, be incoherent to find by the process of backwards tracing that value from the trust fund had been attributed to a certain asset, *prima facie* giving the trust beneficiaries a proprietary stake in it, but then prevent them from taking the benefit of that security in order to protect unsecured creditors. Either the beneficiaries have a proprietary claim, in which case they gain priority over unsecured creditors, or they do not, in which case they do not. After all, we are dealing with 'hard-nosed property rights'.[29] And if it is justifiable for beneficiaries to have priority over unsecured creditors in the context of conventional tracing, it is not clear why a different approach should be taken to backwards tracing: the detriment to the creditors is the same in both cases.

However, this was all that was said on the topic and Lord Toulson did not explain how his policy concerns relate to his test of coordination. Moreover, he did not engage with any of the counter arguments raised above. Therefore, this aspect of his judgement in *Durant* carries little persuasive weight and no firm conclusions can be drawn from it.

---

[24] Conaglen (n 11) 451.
[25] Smith (n 14) 304.
[26] *Smith* (n 14) 304.
[27] *Durant International* (n 4) [13].
[28] ibid [33].
[29] *Foskett* (n 1) (Lord Browne-Wilkinson) 109.

Having said this, admittedly, it is inevitable that if English law were to expand the territory in which proprietary claims can operate, this would sometimes reduce the area in which unsecured creditors can operate. Secondly, this objection that unsecured creditors could be prejudiced carries more weight in the context of backwards tracing than conventional tracing because the law is currently unsettled in this area; for the notion to be officially recognized it would effectively entail a change in the law. This is problematic because many creditors would have entered into arrangements and perhaps declined security, on the presumption that backwards tracing is not possible. It could be regarded as unfair if these creditors subsequently missed out on the entitlements they expected due to a retrospective change in the law.

My response to this argument is simply that this trade-off is necessary because justice nevertheless requires that beneficiaries are able to trace backwards in certain circumstances. This is because, for all intents and purposes, it is *their* money (or other assets) which has been used to acquire the asset in question, albeit indirectly. It is unfair to allow the strict order in which the wrongdoing trustee chose to act to destroy their otherwise sound proprietary claim. By maintaining formalistic, chronological tracing rules, there continues to be arbitrary distinctions relating to the availability of a proprietary remedy, depending only on the particular way the trustee misuses the trust assets. In order to do justice to those parties who have truly been wronged, i.e. the trust beneficiaries, this situation needs to be rectified by the recognition of backwards tracing – even if this would put some unsecured creditors at a greater disadvantage. To address the specific difficulty raised in the previous paragraph, relating to creditors who acted on the basis that backwards tracing does not operate, the courts could potentially invoke the *Re Spectrum Plus*[30] power of prospective common law ruling in appropriate cases in order to preserve their pre-existing rights.

In conclusion, contrary to Conaglen's argument, backwards tracing does not significantly unduly prejudice unsecured creditors, the majority of whom are voluntarily unsecured and therefore run the risk of being displaced in a priority contest on the debtor's insolvency. Moreover, they are not in a radically different position in the context of backwards tracing than they are under conventional tracing rules; any distinction would be illogical and arbitrary because the beneficiaries would have a proprietary claim in both cases. Thirdly, any undue detriment to unsecured creditors, such as where the link between the original debt acquisition and the subsequent use of trust money is very remote, could be prevented by limiting the availability of the backwards tracing process to when certain conditions are met or particular facts exist.[31] Finally, it is accepted that there would inevitably be some negative impact on some creditors if backwards tracing were permitted; however, it is justified in the overall interests of justice and legal coherence.

2. Test for backwards tracing

Having concluded that there should be no absolute bar against recognizing backwards tracing in English law, the next key question to consider is when, and to what extent, backwards tracing should be recognized. It is clear from the preceding discussion that there must be some boundaries on the allowable limits of backwards tracing (for example, to prevent undue prejudice to unsecured creditors where the transactional link is very weak). However, how the test for backwards tracing is conceived depends on what we wish to achieve. In my view, an optimal test would be one which effectively addresses the specific problems identified above, the existence of which justify the expansion of the tracing doctrine in the first place. Furthermore, it should be as certain, clear and easy to apply to the facts of any given case as possible.

The following are a series of different possible tests for determining when backwards tracing is permissible. It is by no means an exhaustive list, but serves to provide a flavour of the different options open to the court, should they decide to recognize the possibility of backwards tracing.

[30] [2005] UKHL 41, [2005] 2 AC 680.
[31] See Section 2, below, for discussion on this.

A. Coordination between depletion of the trust fund and acquisition of the relevant asset

A good place to start is with the test proposed by Lord Toulson in *Durant International*. He states that 'the claimant has to establish a coordination between the depletion of the trust fund and the acquisition of the asset which is the subject of the tracing claim, looking at the whole transaction, such as to warrant the court attributing the value of the interest acquired to the misuse of the trust fund'.[32]

On first reading, this test is logical, flexible and serves to prevent the circumvention of tracing rules by sophisticated money laundering schemes. However, in reality, it is an unhelpful test which essentially imbues judges with an extremely broad discretion to decide the matter in any given case. The problem with Lord Toulson's formulation is that it does not give any principled guidance to the court as to the circumstances where such a 'coordination' will be found to exist; there are clearly cases, such as Burrows' bank account example above, and the case of *Durant International* itself, where the coordination between the transactions is self-evident. On the other hand, there are likely to also be many cases on the margin, where the 'coordination' is less obvious and such a vague test will not help to solve them. For example, it could be said that there is some degree of coordination in any circumstances where misappropriated trust funds are used to pay off a debt which arose on the acquisition of an asset. The statement that the coordination must be 'such as to warrant the court attributing the value of the interest acquired to the misuse of the trust fund'[33] does not provide any solid way of drawing the boundary line.

There is an argument that a wide discretion might initially be necessary to allow the courts some time to flesh out the specific details of the test. However, this does not appear to be what Lord Toulson intended to do in the case. Moreover, such an approach would likely result in inconsistencies and conflicting later judgments in which 'coordination' is interpreted in different ways. In order to best minimise any legal uncertainty caused by recognizing backwards tracing, the test should have a clear structure and realm of application from the very beginning.

On the other hand, reading Lord Toulson's test in conjunction with his comments earlier in the judgement, it becomes apparent that it may actually be too narrow in its scope. For example, as noted above, he stresses the importance of dealing with large scale financial crime and makes specific reference to 'the various steps' in a 'coordinated scheme'. Read in this context, it may be that he meant 'coordination' in his formulation of the test to refer only to coordinated financial schemes, not any form of coordination in the sense of merely closely related transactions. Thus, the objectives of achieving justice for trust beneficiaries whose assets have been used to discharge debts incurred in acquiring other assets and removing arbitrary distinctions from the law of tracing are not well-met by this test. On this view, perhaps Lord Toulson's great leap forward was more of a small step after all.

Whether the latter interpretation is correct or not, Lord Toulson's test is still in need of refinement in order to clarify when backwards tracing is available, if not complete reformulation. Hopefully, this will be provided by the courts in time, fleshing out and building on Lord Toulson's somewhat tentative judgment rather than ignoring it.

B. Temporal test

One possible interpretation of the 'coordination' requirement from *Durant International* is that it refers to some sort of temporal coordination. The court could presume that where transactions occur within a very short space of time of one another, they are sufficiently linked to allow for backwards tracing, regardless of the strict order in which they take place. The burden would then shift on the defendant to demonstrate that, despite the close time connection, there was no such coordination as to warrant tracing. This model explains why Burrows' bank account example works: the credit to the

[32] *Durant International* (n 4) [40].
[33] ibid.

account and the payment in occur almost contemporaneously and it is this fact which justifies allowing tracing into the account even when the payment into the account came after the account was credited. Once the court had decided on the timeframe triggering this presumption, it would become straightforward to apply and the law would be fairly certain.

However, this is nevertheless an unsatisfactory approach, for several reasons. Firstly, it has the potential to be highly arbitrary. There is no simple answer as to where the line should be drawn: transactions which occur on the same day? Within the same week? On the facts of *Durant International*, there was found to be coordination between the payments into the Chanani account which occurred between the 26th January and the 6th February, and the payments from the Chanani account to the Durant account between the 14th and 23rd January. Therefore, clearly some time can elapse between transactions without severing the connection. How much time exactly, however, is a question which invites no easy answer.

One potential solution is that the appropriate timeframe be determined with reference to the context of the case thus allowing for flexibility to deal with different possible fact-patterns. However, this sort of nuanced solution would take significant time to properly work out and its flexibility would come at the expense of simplicity and certainty.

On the other hand, a rigid, temporal approach based on the close proximity of transactions, is unattractive because it fails to take account of the various other ways by which transactions may be connected, for example, to the intentions of the parties involved. It is therefore poorly equipped to deal with the problem of fraudulent schemes that the Privy Council was so concerned about in *Durant International*. It therefore seems clear that Lord Toulson could not have been referring to merely temporal coordination, and rightly so.

C. Fraudulent exception

As explained above, in *Durant International*, the primary, and perhaps decisive, policy consideration was the need to combat fraud. There is a strong feeling that fraudsters should not be able to design their financial transactions with the particular purpose of defeating tracing claims. One solution to this problem would be to maintain the current bar on backwards tracing as a general rule, but carve out an exception for when the court finds that the financial scheme was intentionally structured in such a way as to prevent the operation of tracing.

However, this is too narrow a view. It ignores the possibility that there may be other scenarios, not involving any fraudulent intention, where the transactions so closely linked that it may be desirable to allow backwards tracing.

Moreover, from a practical implementation perspective, this sort of approach could lead the court into some difficult, time-consuming inquiries because trying to identify whether a given financial transaction or scheme is legitimate or a deliberate attempt to circumvent the tracing rules necessarily involves, not only assessing the parties' intentions but also an in-depth investigation of how the scheme operates.

However, courts do routinely use such methods to make these sorts of determinations, so the above cannot be a decisive objection. A stronger argument against such a test is to consider the problems it raises in relation to the relevant standard of proof. Typically, in private law proceedings, disgruntled beneficiaries should be able to vindicate their property rights on a simple balance of probabilities. However, under the Fraud Act 2006, fraud in several different forms is also a criminal offence. This suggests that any finding of fraud should be made on the basis that it is beyond reasonable doubt, a significantly more onerous standard of proof. An analogy could perhaps be drawn with the tort of deceit, for which the claimant must only establish on the balance of probabilities that the defendant made a false statement. However, a finding of fraud could destroy a trustee's

commercial reputation, resulting in dramatic financial consequences for them and thus it is arguably important that such a finding is not made without the due process and safeguards of a criminal trial. This tension highlights the difficulty with a test that involves an investigation into the legitimacy and honesty of the transaction in question, an inquiry that strays dangerously close to criminal law.

D. Test of intention

Another different, and, it is submitted, better, way of defining the requisite coordination is by reference to the intention of the trustee(s) who misappropriate(s) the trust assets.

Returning to the case of *Foskett*, in the Court of Appeal, Sir Richard Scott[34] was a strong advocate of permitting backwards tracing, commenting that there is no reason 'why the order in which the events happen should be regarded as critical to the claim'. However he does not propose that backwards tracing should always be allowed; the limit he seems to suggest is that it must be shown that 'it was always the intention to use the trust money to acquire the asset'.

More recently, Virgo and Davies[35] have similarly suggested that 'where a defendant incurred a debt in order to acquire an asset, and always planned to discharge the debt with a beneficiary's property, then this seems likely to satisfy the "coordination" and "close causal and transactional link" requirements demanded by the Privy Council'.

Employing the concept of intention as a limiting device has been suggested by other proponents of backwards tracing, and I argue that they are correct to do so. This is because it provides a sufficiently strong and ascertainable link between the original transaction, by which the asset in question is acquired, and the second transaction, the discharge of the debt incurred in the first. The intention for the trust money to 'take the place' of the debt means that the asset acquired by the incurring of the debt can be regarded as the genuine substitute asset for the trust money. This is consistent with Smith's proposition that the money used to discharge the debt constitutes 'delayed payment'. It is this intention, linking the relevant transactions, which warrants allowing backwards tracing.

The flipside of this argument is that it would be completely illogical to allow tracing into an asset purchased on credit where it was never intended that the trust money be used to discharge the debt incurred. For example, when somebody incurs a debt when acquiring an asset before they are even appointed to be a trustee, and then later use misappropriated trust funds to pay it off, there is less strong justification for allowing the beneficiaries of the trust to trace into that previously acquired asset: the acquisition and the discharge of the debt might be two entirely detached transactions if there was no intention at the time of acquisition that the trust money would be used in any way to fund it. The fact that it was so used is purely coincidental and hence insufficient to justify backwards tracing. Similarly, there would be no backwards tracing where trust funds were mistakenly used to pay off a previously incurred debt because the necessary intention is missing. Just as it is the presence of such an intention which sometimes links two transactions, it is the lack of intention which separates them.

This appears to be the logic relied upon by Virgo and Davies; they argue that 'where a defendant incurred a debt in order to acquire an asset, and always planned to discharge the debt with a beneficiary's property', 'allowing the beneficiary to assert a proprietary interest in the acquired asset might not unfairly prejudice the defendant or other creditors since the asset would never have been acquired by the defendant had he or she not known that he or she would be able to exploit the

---

[34] *Foskett* (CA) (n 7), 283.

[35] Graham Virgo and Paul Davies, 'Backward tracing' (*OUPblog*, 17 May 2016) <http://blog.oup.com/2016/05/backward-tracing-law/> accessed 1 October 2016.

beneficiary's property'. They introduce a requirement of a finding that the asset would not have been acquired or the debt not incurred *but-for* the intention that trust assets would be used to pay for it. This certainly provides a substantial link between the two transactions. However, it is a requirement which arguably limits the scope of Virgo and Davies' test for backwards tracing to too great an extent, failing to cater for other fact-patterns which may also warrant backwards tracing. One example would be where a trustee agrees to take on that role solely because he wants to have access to trust funds to pay off previously incurred debt. In this scenario, there may have been no intention to use misappropriated trust funds to pay the debt at the moment the debt was incurred; however, there was such an intention when the trustee took up their trusteeship. It is submitted that, contrary to the view of Virgo and Davies, this latter intention should also be sufficient for backwards tracing to operate; it constitutes a sufficient link between the two transactions because the trusteeship was only taken due to an immediate desire to use the newly acquired funds to pay off the debt.

Thus the test should be formulated more precisely as allowing backwards tracing when a trustee:

1. Misappropriates trust assets to pay off an existing debt; and
2. Had an intention or desire to use trust assets to do so, either
    a. At the time the debt was originally incurred; or
    b. At the point the trusteeship was taken up.

Fundamentally, this two-fold formulation of the intention based test seeks to identify a real link between the transaction which incurs the debt and the later use if trust funds to discharge it. This is why it is not enough that trust funds are mistakenly used to pay off a previous debt.

A possible objection to this test is that it focuses solely on the intention of the trustee even though there are likely to be cases where the trustee is insolvent and the suit is therefore only between the creditor and beneficiary, whose rights then depend upon the actions of a third party whom neither has control over. However, this position is justified because one of the most pressing justifications for permitting backwards tracing, as explored above, is preventing the trustee from succeeding in a fraudulent scheme. Moreover, on a conceptual level, it is only the intention of the trustee when dealing with the trust assets which can legitimately indicate whether or not the property in question can properly be regarded as a 'substitute' for the misappropriated trust assets (the inquiry upon which the whole process of tracing, backwards or otherwise, hinges); the intentions of either the creditors or beneficiaries are irrelevant in this regard.

Furthermore, this intention based test is preferable to Lord Toulson's formulation in *Durant International* because it does not rely on the court's discretion: the requisite transactional link, or 'coordination', will *always* be found where the requisite intention is found.

In his article, Conaglen floats this idea of limiting the scope of backwards tracing by reference to the trustee's intention at the time the asset was acquired. However, he discounts it on the basis of the 'evidential difficulties inherent in a test that is focused on the defalcating trustee's intentions'.[36] In *Durant International*, Lord Toulson similarly did not seem attracted by the idea of an intention-based test, despite its previous judicial endorsement. An insight into why this is so can be gained from his statement that the establishment of the requisite coordination is 'likely to depend on inference from the proved facts, particularly since in many cases the testimony of the trustee, if available, will be of little value'.[37] He, like Conaglen, doubts the utility of an intention-based test in practice.

It is perhaps true that the actual personal testimony of a trustee who has misused trust assets may not be the most reliable. However, this counter argument is weak because judges frequently employ tests of intention in all areas of the law, including equity. The personal liability of third parties

[36] Conaglen (n 11) 455.
[37] *Durant International* (n 4) [40].

to breaches of trust is a good example: to prove dishonest assistance or knowing receipt, the intention of the assistor or receiver is crucial and accordingly, the courts have to determine this issue. Admittedly, there can be evidential uncertainty in such tests, especially when they are subjective. However, these difficulties can be circumvented by adopting a quasi-objective test, based, not on what the parties involved actually did intend, but on what a reasonable person would infer the parties' intentions to be from their actions and all the circumstances. Judges are experienced at taking into account all of the evidence available to them in their inquiries into intention. The application of an intention based test for backwards tracing would be no different.

Nevertheless, there will inevitably be difficult, complicated fact-patterns which are not clearly covered by this relatively simple formulation of the intention test. One example would be where the trustee originally intended to pay off only one debt with the trust assets, but ended up paying off several others as well; the original debt would clearly be captured by the test but the other payments would not fit as obviously. Or where the trustee intended to pay off Debt A but later decided to pay off Debt B instead. These hard cases inevitably will have to be resolved sensibly by the courts, keeping in mind the goals they seek to achieve by permitting backwards tracing in the first place (namely preventing financial crime and doing justice between the parties). For example, one possible solution to the second hard case described above would be if the court decided that an intention to pay off *any* debt at the point of taking the trusteeship is sufficient even if the specific debt subsequently paid was not envisaged at that time. Alternatively, the court may hold that backwards tracing only operates where the intention is to pay the particular debt which is paid off; on this view, a general intention to discharge debts is not enough. It is submitted that this latter approach is preferable because it better preserves a strong link between the incurring and subsequent discharge of the debt, and prevents backwards tracing from operating too broadly, in situations where the relevant assets cannot sensibly be described as substitutes for the misappropriated trust assets (which would be the case where the intention, i.e. to pay Debt A, did not strictly match the later action, i.e. paying Debt B).

3. *Conclusion*

The law, like the commerce it regulates, is a living, breathing, ever-evolving entity. As a result, laws which were once justifiable and appropriate can later appear outdated and in need of reformulation. The current embargo on backwards tracing in English law is just such an example. The world in which we live has changed dramatically since the days where transactions were carried out with gold bullions in a bag;[38] very many financial transactions take place using online systems, and they can be extremely complex, involving several countries, multiple currencies and a substantial period of time.

This is the context in which *Durant International* was decided; Lord Toulson was very much alive to the risks of high value financial crime and this was one of the main reasons why he was willing to recognize that backwards tracing is indeed the only way forward. This was an important step which hopefully the English courts will follow. However, the English courts should not follow the Privy Council in adopting the vague, yet limiting, proposed test of coordination to determine when backwards tracing should be permitted. Instead, the most practical and conceptually sound way of implementing backwards tracing is by reference to the intentions of the parties, ensuring that cases where backwards tracing is allowed are restricted to where there has always been an intention to use the funds held on trust to make the transfer or acquire the asset. This solution is just, conceptually coherent and brings the law of tracing in line with modern day commerce, where debt plays a key role. If it were ultimately adopted, the case of *Durant International* would undoubtedly have played its part in getting us there, although maybe not for entirely the right reasons!

[38] *Taylor v Plumer* [1815] EWHC KB J84 (CA).