Aberdeen Student Law Review

Volume 11

2022

Kindly sponsored by

EDITORIAL BOARD
VOLUME 11

FOREWORD BY THE RT HON LADY DORRIAN LORD JUSTICE CLERK AND PRESIDENT OF THE SECOND DIVISION OF THE INNER HOUSE OF THE COURT OF SESSION

I am very pleased to introduce the new volume of the Aberdeen Student Law Review. The Review continues to engage with the pressing legal issues of the day. Topics of current and sensitive relevance include the use of jury trials, assisted dying and oil and gas decommissioning, all subjects of enormous public interest and importance. The list of topics in this issue alone is testimony to the breadth of student interests at Aberdeen and to the strength of their writing. I hope you have a good read!

Rt Hon Leeona Dorrian
Edinburgh, March 2022

NOTE BY DR DOUGLAS BAIN
SENIOR LECTURER AND CONTRIBUTOR TO VOLUME 3

I remember it as being around spring 2010. Dominic Scullion and I were sitting on a bench in the cloister outside Elphinstone Hall (west-facing) and it was, from the position of the sun, as remembered, late afternoon. I mention the sun because of the light, which I particularly recollect; Old Aberdeen is always at its best in the spring and the autumn, on account of that light. Anyhow, Dominic was telling me about an exciting new project for a refereed law journal that had just been green-lit by the (then) Head of School, Professor Margaret Ross: the self-same journal the eleventh issue of which you currently hold in your hands (literally or metaphorically). Dominic was, at that time, nearing the end of his Diploma in Legal Practice, and I was desperately wishing I could begin to see the end of my PhD.

Much water has flown beneath the bridge since that afternoon.

The journal, as conceived, was to be a home for a wide range of articles on a broad range of subjects, from undergraduate essays to honours dissertations and papers from doctoral candidates. It was also to include articles and notes from members of Law School staff and from alumni. My own article, co-written with a fellow alumnus, is in volume 3. I'll have to have a go at writing another sometime, but the worry is that so stiff is the competition these days, I'm not sure it would make the cut. I say this with some certainty having been an avid reader of the journal – now available, please note, via HeinOnline

and Westlaw, in addition to through the University of Aberdeen Law School's web pages – since volume 1, and also having had some oversight of the pool of submissions for the issues.

As regards that pool of submissions, looking at the list of accepted authors, I see former students who have gone on to complete higher degrees, some of whom are now on staff at Aberdeen University Law School, or in other law schools. I have found authors whose first publication credit was in ASLR who have gone on to be published in, for example, Juridical Review, the Journal of the Law Society of Scotland, and Edinburgh Law Review. I see trainee solicitors, admitted solicitors and advocates; and not just on the list of authors, but also the annual lists of editorial teams. When will ASLR net its first former contributor Lord President? I can't say, but it's got to be on the cards.

Most rewarding for me are the published articles that have been reworked from law honours dissertations. I was for six years, to 2019-20, the law honours dissertation coordinator. Each year, one way or another, I would have seen or exchanged emails with just about every student in that year's dissertation cohort. I can see in past ASLR tables of contents a number of articles which had their genesis in dissertations for which I was the student's staff consultee.

Each year, at the instigation of the journal, I would send an e-mail to the graduating class drawing the students' attention to the Aberdeen

Student Law Review as a possible host for the wider circulation of their work. In recent years, until Covid intervened, I co-hosted, along with the journal editors, a graduation day breakfast meeting, at which the possibility of reworking for publication was presented and options for publication discussed.

Of course, ASLR, though an early pioneer, is not the only such student-led law journal in Scotland or beyond. There are many others. Nor is it the only entry route into publication in the University of Aberdeen law community. The Law School has, in its web pages, since October 2015, hosted an active law blog,[1] which frequently includes contributions from students.

There is a quote, sometimes attributed to Brian Eno, about the band Velvet Underground: 'The first Velvet Underground album only sold 10,000 copies, but everyone who bought it formed a band.' Music and football fanzines have, historically, had similar galvanizing effects. In terms of the study and practice of law, ASLR has, by providing an example and template, and a gateway into formal, refereed, widely-disseminated and accessible publication, provided students with a platform to show off their work – the fruits of their law studies – and also useful experience, a publication credit, an opportunity for networking, something for the CV, etc.

[1] www.abdn.ac.uk/law/blog

It also provides something for friends and family to see. Not to mention being a capstone on their law studies and a launching platform for their future in law.

None of this overstates the case.

Dr Douglas Bain
Aberdeen, August 2021

EDITORIAL FOREWORD

We are indebted to the School of Law for their unwavering support. In particular, thanks are due to Dr Alisdair MacPherson, our Honorary Secretary, without whom this volume would not have been possible, all those who have peer reviewed papers, and Georgi Chichkov and Carol Lawie in the Law School office. Gratitude is also due to our sponsor, Stronachs, for their continued interest in the journal.

This has been a hugely successful year for ASLR, with a record number of submissions. We owe great thanks to all those who submitted articles. The quality of submissions was excellent; the challenge of selecting articles for publication immense.

Our editorial team have been indispensable in helping us manage this workload. To them, we are very grateful. Due to the pandemic, our meetings have been confined to video calls. However, this has also demonstrated the international reach of ASLR, with members of our team based as far afield as Canada and South Africa.

This year's volume is significant in its making, as all submissions were written, edited and reviewed amidst the backdrop of the pandemic. Historically, isolation has been a great source of creativity. Great works of literature such as Shakespeare's King Lear were written during isolation, as well as famous pieces from the art world such as

van Gogh's Starry Night. We hope that the readers of this review see a small glimmer of this creativity borne out of unique circumstances.

Without further ado, we are delighted to introduce volume 11 of the Aberdeen Student Law Review.

Robbie Callander
Sheyda Rimmer
Edinburgh, May 2022

TABLE OF CONTENTS

Articles

Case Notes

Let the Layman Lie: The Case for the Abolition of Jury Trials in Scotland

JOHN LOGAN*

Abstract

The use of juries in criminal trials in Scotland is deeply ingrained in the public consciousness. This article will discuss the efficacy of their use with particular reference to the substantial issues surrounding jurors' competence, prejudice, and media-imparted bias when serving in criminal cases. It will go on to address potential reforms to the jury system that may alleviate some of these issues. A comparison will be drawn with South Africa, where the use of juries has been eliminated altogether. The article will conclude with the recommendation that, on the basis that the Scottish jury system is beyond repair, the interests of justice would be best served by the abolition of the jury in Scottish criminal trials.

Keywords: juries, law reform, Scots criminal law, Scots procedural law, South African criminal law

Introduction

For centuries, juries have been a facet of Scots law. While the exact moment of their conception has been lost to the ages, they can be traced back as far as the Norman infiltration of Scotland in the twelfth century.[1] The notion that average citizens, untrained in the legal profession, should have a role in court as the trier of facts has been firmly established as a cornerstone of our legal system. In practice today, the presence of the Scottish jury is dependent on whether the case is one involving summary or solemn proceedings. Summary proceedings involve the lower courts, and do not give rise to a jury.[2] Solemn proceedings, on the other hand, involve the higher courts, and

* LLB (Hons), 2020; DPLP, 2021, University of Aberdeen.

[1] I D Willock, *The Origin and Development of the Jury in Scotland* (Stair Society 1966).

[2] Peter Duff, 'The Scottish Criminal Jury: A Very Peculiar Institution' (1999) 62 LCP 173, 175.

constitute both a judge and a jury.[3] The latter has exclusive jurisdiction over the most serious common law crimes of rape and murder; that is to say, these offences are to be tried as solemn proceedings, in the presence of a jury. Conversely, a number of statutory offences are to be tried as summary offences, without a jury.[4] In practice, however, the majority of offences are be considered 'either way' offences, and, as such, may be tried with or without a jury. This depends on the course the Crown elects to take with the charge, considering, *inter alia*, the seriousness of the case specifics.[5] A minor assault or theft would be more likely to result in summary proceedings, whereas a more violent assault or grander theft would be more likely to result in solemn proceedings.

This is of particular note for its stark comparison to the situation in England. The English defendant may elect whether or not to be tried by a jury, whereas the Scottish accused has no such choice. Further, given that juries are only engaged in trials dealing with the most serious offences, it can be said that the jury is involved when the need for the effective application of the law is most crucial. It is therefore of the utmost importance that they are shown to be just, fair, and effective tools of the court whenever they are summoned into existence. The longstanding legal tradition of involving laypersons in the trial process has received much support from lawyers, academics, and the public. It is seen as a balancing factor, limiting the power of the state through its inclusion of the ordinary public in a decisive role in the practical application of the law. Indeed, Lord Devlin went so far as to praise the jury trial as 'the lamp that shows that freedom lives'.[6] With such vehement support for juries across stakeholder groups, it is not hard to see why they have remained a stalwart of a greatly changing system over the centuries. The jury trial has even been exported across the globe, constituting an aspect of multiple international legal systems.[7]

---

[3] ibid.
[4] ibid 176.
[5] ibid 177.
[6] Patrick Devlin, *Trial by Jury* (Stevens & Sons 1956) 164.
[7] Valerie P Hans, 'Jury Systems Around The World' (2008) 4 Annual Review of Law and Social Science 305.

Despite the widespread general popularity and support for juries in both the law and the media, they are not without their detractors, who criticise aspects of our current jury system, or indeed the concept of a jury as a whole. Darbyshire, for example, directly attacks Devlin's stirring sentiment, deriding the public attitude towards juries as merely an indefensible sentimental attachment to their symbolic function.[8] Criticisms regarding the competence of the jurors themselves are also common; an American judge, Jerome Frank, remarking of them that:

> While the jury can contribute nothing of value so far as the law is concerned, it has infinite capacity for mischief, for twelve men can easily misunderstand more law in a minute than the judge can explain in an hour.[9]

Closer to home, Callander goes so far as to call for the full abolition of juries in rape trials, citing commonplace prejudicial attitudes of jurors towards victims in cases of rape.[10] Further, government reviews throughout the UK have given great consideration to the efficacy of juries on multiple occasions. The Auld Report of 2001 observed that judges should have the right to try serious fraud cases with the assistance of expert lay assessors, rather than jurors.[11] Even more radically, the report sought to remove the current option for those in England accused of some midlevel offences, such as theft, assault, or drug offences, to elect whether to be tried by a jury, instead leaving that decision up to the magistrate.[12] The latter suggestion is less relevant to Scots law due to some specific aspects inherent within it, but not its English equivalent, which will be discussed anon. Nonetheless, it is still telling that there is a degree of suspicion surrounding juries within the professional world. More recently, the Scottish Government conducted the largest mock jury

[8] Penny Darbyshire, 'The Lamp That Shows That Freedom Lives—Is It Worth the Candle?' (1991) Crim LR 740.
[9] *Skidmore v. Baltimore* 167 F2d 54 (2d Cir 1948).
[10] Isla Callander, 'The Jury Is An Inappropriate Decision-Making Body In Rape Trials In Scotland: Not Guilty, Not Proven, Guilty?' (LLM thesis, University of Glasgow 2013).
[11] Robin Auld, *Review Of The Criminal Courts Of England And Wales* (Stationery Office 2001).
[12] ibid.

study ever undertaken in the UK, in order to better understand the nature of juries in the unique Scottish system, and by extension, their efficacy in law.[13] Such an exertion of time and cost would likely not have been undertaken were the view towards juries that of contented acceptance.

It seems then that this rosy picture of what has long been an accepted and mostly unchallenged aspect of the criminal justice system is not universally held. There are certain dissenting notions that call into question the efficacy of the jury system as a whole. Indeed, juries in civil cases were officially abolished by legislation in the Sheriff Courts in the 1980s,[14] and while they are still theoretically allowed in civil trials in higher courts, such cases are extremely rare.[15] This article serves to discuss the use of juries specifically in criminal cases in Scotland; determining whether they are in fact an effective tool of justice, or whether the principles of truth and fairness, which must be inherent in any criminal justice system, would be better served by way of their abolition. Chapter one will illustrate some of the issues which plague the current use of jury trials in Scotland; chapter two will discuss if these issues could be overcome with reform of the current jury system, shy of abolition; and chapter three considers what potential replacement could better serve the application of justice in criminal trials, were juries to be abolished.

Issues with the Jury

Juror Competence

Perhaps one of the most recurring considerations regarding the merits of the jury system is the question of whether a group of randomly elected lay individuals, untrained in the field of law, have the competence to understand the law and evidence, and their application in the case at hand. Critics make reference to the convoluted intricacies that the legal field is infamous for, with technicalities, exceptions, and complicated rules abounding. These can often be difficult to

[13] Rachel Ormston and others, *Scottish Jury Research: Summary Of Findings From A Mock Jury Study* (Scottish Government 2019).
[14] Law Reform (Miscellaneous Provisions) (Scotland) Act 1980, s 11.
[15] Duff (n 2), 174.

understand fully without the proper legal education. The aforementioned Judge Frank colourfully argued that lay jurors would be as unable to understand the legal words of the judge as if they were spoken in Chinese, Sanskrit, or Choctow.[16] Conversely, proponents of the jury trial claim that the combined wisdom of multiple heads is superior to that of one, even if that head is attached to a learned professional judge.[17]

A difficulty arises in that the verdict alone may not be a sign of jury competence, or lack thereof. The very nature of an adversarial trial is that it involves two opposing sides peddling their incompatible truths as fact; the decision of which amounts to the genuine truth left up to the trier of fact themselves, be that judge or jury. The fact that jurors return a verdict for conviction or acquittal could well be the result of the jurors simply being better convinced one way, rather than their decision being based on an understanding of, or lack thereof, the evidence or a point of law. As such, it is necessary to establish some yardstick by which juror competence can be measured. Bornstein and Greene propose that one potential measure of juror competence is their comprehension of the instructions presented to them by the judge, intended to assist them in producing a legally sound verdict.[18] Unfortunately, mock juror research has found that jurors' comprehension of these instructions are consistently disappointing, falling as low as a mere 50%.[19] Comiskey notes, with regards the views of researchers into jurors' comprehension of legal tests, they have:

> [A]lmost unanimously concluded that a jury's ability to comprehend [them] is poor and that there is room for considerable improvement.[20]

---

[16] Jerome Frank, *Courts On Trial* (1st edn, Princeton UP 1973).
[17] Phoebe C Ellsworth, 'Are Twelve Heads Better Than One?' (1989) 52 LCP 205.
[18] Brian H Bornstein and Edie Greene, 'Jury Decision Making: Implications For And From Psychology' (2011) 20 Current Directions in Psychological Science 63.
[19] Chantelle M Baguley, Blake M McKimmie and Barbara M Masser, 'Re-Evaluating How To Measure Jurors' Comprehension And Application Of Jury Instructions' (2020) 26 Psychology, Crime & Law 53.
[20] Marie Comiskey, 'Initiating dialogue about jury comprehension of legal concepts: can the 'stagnant pool' be revitalised?' (2010) 35 Queen's LJ 625, 629.

While concerns have been raised regarding the effectiveness of the means by which juror comprehension is studied,[21] these repeatedly poor results do not fill one with confidence as to the competence of jurors to apply complicated law. If they struggle to understand the instructions intended to ensure their verdict is legally compatible, then doubt is raised as to whether they can be trusted to return a legally sound verdict at all.

Furthermore, issues of jurors' difficulties in understanding complicated legality are of particular note when applied in a Scottish setting. In 2018, a Scottish jury research study involved 64 mock juries who were asked to reach a verdict in a fictional trial, concerning either a rape or a physical assault.[22] Within the findings, the researchers recognised a frequency of misunderstandings of the key Scottish concepts of corroboration and the not proven verdict. With regards the former, it was often the case that the mock jurors believed, mistakenly, that the specific forensic evidence in the rape trial must demonstrate unequivocally that a rape had occurred and that this was corroborated, allowing the jury to convict.[23] With regards the not proven verdict, the study also highlighted an example of a crucial misunderstanding on the part of a juror who, again wrongly, believed that the accused may be later retried by the courts in the case of a not proven verdict.[24]

Clearly there exists a potentially dangerous limit to jurors' competence, particularly in the high stakes of a criminal trial. Research has shown that untrained individuals struggle consistently to understand instructions posed to them by judges, or complicated legal concepts. It is not difficult to speculate from the examples above a jury convicting an accused based on the erroneous belief that the corroboration requirement had been met; or that by acquitting under the not proven verdict, the accused may be retried for the same offence at a later date. While it would be foolhardy to believe professional judges to be omniscient of all legal knowledge, one would hardly

[21] ibid.
[22] Rachel Ormston and others, *Scottish Jury Research: Summary Of Findings From A Mock Jury Study* (Scottish Government 2019).
[23] ibid.
[24] ibid.

expect them, with their years of legal education and experience, to make such fundamental errors in understanding of legal rules and concepts.

Jury Secrecy

A related problem is that of the secrecy of jury deliberations. It is enshrined in statute that it is a contempt of court to seek or disclose any particulars put forth during the deliberations of a jury in any legal proceedings.[25] This means that the specific reasoning by which a jury reaches its decision may not be known by anyone outwith the bounds of the jury room itself. This is in contrast to professional judges as, whilst there is no statutory requirement for judges to explain their decisions as there is in some continental jurisdictions, case law has reflected that the absence of this explanation may itself be good grounds for an appeal.[26]

There are valid arguments for allowing confidentiality of the discussion of the jury in the jury room. It allows for free and uninhibited discussion among jurors, and protects them from harassment should they be known to have spoken out against the unsavoury characters which are an inescapable aspect of criminal trials. However, this confidentiality carries with it a dangerous lack of accountability given that none outside the jury room will ever know how the decision was reached, whether by legally and logically sound reasoning or otherwise. One infamous case which highlights the danger of confidential and unchecked jury discussions is that of *R v Young*.[27] The case involved the defendant, Stephen Young, being accused of committing a double murder, of which the jury found him guilty.[28] However, it was later revealed that the foreman and three other jurors had conducted a séance at the hotel they were staying at with the use of a Ouija board, which supposedly pointed to Young as the guilty party.[29] It is unlikely controversial to state that the jurors in

[25] Contempt of Court Act 1981, s 8.
[26] J R Spencer, 'Inscrutable Verdicts, the Duty to Give Reasons and Article 6 of the European Convention on Human Rights' (2001) 1 Archbold News 5.
[27] *R v Young (Stephen)* [1995] 2 WLR.
[28] ibid.
[29] J R Spencer, 'Seances, And The Secrecy Of The Jury–Room' (1995) 54 CLJ 519, 520.

their deliberations should not consider evidence gathered by way of the occult. In the event, the conviction was quashed at the Court of Appeal for this reason, although Young was later retried and again found guilty.[30] This is a particularly unique and shocking example of deliberations of jurors not being based in either legality or logic; though it nevertheless highlights the general risk of jury confidentiality. Indeed, the court considered that it was only by luck that the séance took place at the hotel, and not as part of the deliberations within the jury room, in which case the séance would not have been discovered.[31] This issue was considered in the later case of *R v Mirza*,[32] where it was held that the judge would be right to investigate events within the jury room where the deliberations could not even be described as such. In that case, Lord Craighead gave examples of reaching a verdict by the drawing of lots, or by simply flipping a coin.[33] Nonetheless, this is a rare exception to the general rule of jury confidentiality and it was specifically noted in *Mirza* that inquiry is not required in every case, stressing the discretion of the sitting judge.[34]

The reluctance of judges to scrutinise the confidentiality protection afforded to juries may open the door to questionable non-legal reasoning. Whilst *Young* is a particularly stark example, there are many lesser shades of potential mischief which may still amount to wrongful deliberations taking place within the walls of the jury room. One could hardly imagine a professional judge, in his explanation for reaching a verdict, claiming that he was informed of the accused's guilt or innocence by way of consultation with an Ouija board.

Prejudice of the Jury

The previous discussions regarding both the potential for a lack of understanding on the part of jurors, and foolish or illogical reasoning and deliberation, illustrate some of the issues of involving lay juries in the criminal justice process. However, while the risks inherent to jury

[30] ibid.
[31] *R v Young* [1995] QB 324.
[32] *R v Mirza* [2004] UKHL 2.
[33] ibid.
[34] ibid.

trials should not be discounted, these issues often arise in good faith. In general, jurors are not equipped to comprehend complicated legal concepts and arguments; though there is no actual malice, subconscious or otherwise, in their actions.

A darker issue is the risk of prejudicial biases held by jurors, even before they step into the courtroom. Such views represent a direct assault on the crucial pillars of justice, impartiality and fairness, which are inherent in any well-functioning legal system. A conviction or an acquittal must be reasoned upon the legally admissible evidence presented in the trial, and this alone. Nonetheless, the random nature of jury selection always allows for the possibility of certain prejudiced individuals to find themselves on a jury, wielding life-changing power in their decision to convict or acquit. Such biases may be along racial lines regarding the accused;[35] they may arise from prejudicial social attitudes, which is a particular problem in cases of sexual assault;[36] or they may arise from publicity of the trial created by a media circus.[37]

Personal Prejudice

It is a sad fact that prejudicial hatred is prevalent to at least some degree in society. Recent statistics from the Crown Office and Procurator Fiscal Service (COPFS) found that, while there has been a decrease in recent years, there were still almost three thousand charges of racially aimed hate crimes in Scotland between 2018 and 2019,[38] and over five hundred religiously aggravated charges.[39] There is therefore a strong possibility that an individual who holds strong prejudices may, at some point, find himself serving on a jury. This is inherently problematic for the effective application of the law, where impartiality is crucially important, to say nothing of the duties of the state under Article 6 of the European Convention on Human Rights to

[35] Gillian Daly and Rosemary Pattenden, 'Racial Bias and the English Criminal Trial Jury' (2005) 64 CLJ 678.
[36] Callander (n 10).
[37] Lorraine Hope, Amina Memon and Peter McGeorge, 'Understanding Pretrial Publicity: Predecisional Distortion Of Evidence By Mock Jurors.' (2004) 10 Journal of Experimental Psychology: Applied 111.
[38] *Hate crime in Scotland 2018-19* (COPFS 2019).
[39] ibid.

ensure an individual's right to a fair hearing.[40] Such rights may be infringed upon if the fate of the accused is left in the hands of one who holds prejudicial views of hatred towards them for a characteristic wholly independent from the indictment.

Such fears have been shown not to be baseless. They were at issue in the aforementioned case of *R v Mirza*,[41] in which the House of Lords dealt with racial bias in the jury. The facts involved a Pakistani man, Mirza, being found guilty by a jury for charges relating to indecent assault. It was later alleged by one of the jurors that some of the other jurors were racially prejudiced against Mirza and were not impartial. However, the court held that, referencing the decision in *R v Qureshi*,[42] itself involving similar facts of allegedly racist jurors, statements intrinsic to the deliberations of the jury were inadmissible, insofar as they were revealed after the verdict had been delivered, even when alleged bias was claimed.[43] In other words, the courts were powerless to investigate whether the alleged racial biases of jurors within the *Mirza* trial had any basis in reality or whether the guilty verdict reached was in any way caused by them.

Defenders of the jury system have claimed that there exists workable safeguards within the Scottish jury system to avoid personal prejudice clouding the application of justice. In *McCadden v HMA*,[44] the accused, who was Catholic, appealed his conviction on the basis that one of the jurors, a Protestant, was overheard bragging about his plans to find McCadden guilty and making sectarian comments about him. However, Lord Wheatley argued that the wide pool and randomness of jury selection should serve to balance out any extreme prejudices within the jurors; that peremptory challenges allows for the defence to exclude certain jurors without reason; and that one should not 'lightly assume' that jurors would both betray their oath as jurors and ignore directions from the judge to put aside their own preconceived notions in determining the outcome of the case.[45] This

[40] Convention for the Protection of Human Rights and Fundamental Freedoms (European Convention on Human Rights, as amended) (ECHR) art 6.
[41] *R v Mirza* [2003] UKHL 2.
[42] *R v Qureshi* (Sajid) [2001] EWCA Crim 1807.
[43] *R v Mirza* [2004] UKHL 2.
[44] *McCadden v HM Advocate* 1985 JC 98.
[45] ibid.

argument does not stand up to scrutiny. The mere fact that jurors are randomly selected from the electoral register does not allow for the inference that any prejudicial view would be balanced out in the mix of fifteen, particularly given the fact that juries in Scotland may convict on a simple majority. As such, the odds at which a prejudicial juror could have an active effect on the case from his own views drops to a mere one in eight. Further, Wheatley's reference to peremptory challenges as a potential safeguard, whilst a valid point in 1985 when the case was heard, no longer applies as the use of peremptory challenges was abolished in Scotland in 1995.[46] Finally, the notion that we cannot assume that jurors would not put their prejudices aside when deliberating, or that the judge instructing them to do so amounts to sufficient steps to overcome that prejudice, seems to be overly hopeful and even cavalier, given the high stakes of verdicts in jury trials. It seems, then, that the current jury system in Scotland is not properly equipped to deal with the risk of jurors' personal prejudice.

This is not to say that prejudice plays a part in all, or even any sizeable number, of jury trials. It merely demonstrates that there exists a genuine risk, which may very well have already transpired in multiple cases, of biases negatively affecting the equal and impartial application of justice in jury trials. This is compounded by the secrecy surrounding such deliberations. One may go so far as to say that the guaranteed right to a fair trial, devoid of personal discrimination, is being sacrificed on the altar of the jury.

Prejudice in Rape Cases

In addition to the risk of prejudice towards the accused, it is also the case that prejudices may occur with regards the offence itself or the victim. This is of particular significance for sexual offences, as these crimes tend to take place in private. For jurors at trial, there is often great emphasis on the credibility of a victim's testimony. It is very often resorted to a case of 'he said, she said'. As such, there is a real risk regarding the credibility of a rape victim by way of extrajudicial prejudice, given the predilection of jurors towards believing 'rape myths'. In 2018, the Labour MP Ann Coffey set out some of the

[46] Criminal Justice (Scotland) Act 1995, s 8.

pervasive myths believed by jurors during parliamentary debate. She noted that research has demonstrated a belief in society, and thus juries, that rape victims may be to blame by way of their consumption of alcohol, their manner of dress, or their flirting with the accused.[47] Further, she notes that jurors may not believe that the act constituted "real rape" if the victim did not attempt to fight off her attacker, sustain injuries, or report the assault immediately.[48] This is despite the fact that the legal requirement of force to constitute non-consensual penetration as rape was removed by *Lord Advocate's Reference No.1 of 2001*;[49] that it is often the case that victims of rape do not physically resist their attacker or sustain injuries from him;[50] and relatively few victims report their rape within the immediate period after the event, with many waiting weeks, months, or even years.[51] The mere fact that a particular rape does not fit into the narrow confines of what is classically considered to be rape does not lessen the destructive effect it has on the victim, nor does it make the offender any less guilty in law. Nonetheless, mock jury research has shown that these myths are pervasive in the minds of jurors, and can have real effects on their deliberations and final outcome of the trial.

In 2009, Ellison and Munro conducted extensive research into the effect of mock juror attitudes in rape trials, with shocking results showing a consistent approach of jurors to disbelieve a victim's claim of rape, resulting specifically from ingrained beliefs in rape myths.[52] Individuals in the study presented with a mock victim who had suffered no physical injury consistently considered this lack of injury was significant to their delivery of a not-guilty verdict.[53] This was despite presented evidence of a lack of any positive signs of consent,

---

[47] HC Deb 21 November 2018, vol 649, cols 345-346.
[48] ibid.
[49] Lord Advocate's Reference No 1 of 2001 2002 SLT 466.
[50] Sandy Brindley and Michele Burman, 'Meeting the challenge? Responding to rape in Scotland' in Nicole Westmarland and Geetanjali Ganjoli (eds), *International Approaches to Rape* (Policy Press 2012).
[51] Stephanie Lanthier, Janice Du Mont and Robin Mason, 'Responding To Delayed Disclosure Of Sexual Assault In Health Settings: A Systematic Review' (2016) 19 Trauma, Violence, & Abuse 251.
[52] Louise Ellison and Vanessa E Munro, 'Reacting To Rape: Exploring Mock Jurors' Assessments Of Complainant Credibility' (2008) 49 British Journal of Criminology 202.
[53] ibid 205.

and even verbal resistance.[54] Further, even when the victim did present with physical injuries, the jurors seemed quick to dismiss these as the wrong type or lacking in intensity enough to convince them of the victim's version of events, with many going so far as to concoct outlandish alternative means by which the victim could have sustained her injuries. These included her wearing a hard necklace, riding a rollercoaster the day after the event, or even self-inflicting them to better support her claim of rape.[55]

Doubts regarding credibility arising from a delay in the reporting of the rape were also recurring. While some mock jurors did acknowledge that it may be understandable for a rape victim to not immediately turn to the police, others considered a delayed response to be damaging to the prosecution's case, with many expressing credibility concerns as to why the offence was not reported right away.[56] Conversely, it was also found that the opposite could be considered problematic by the jury, with some jurors expressing disbelief that an individual who had suffered rape would so quickly turn to the police after such a traumatic experience, or before contacting family or friends.[57] It seems there exists a case of "damned if you do, damned if you don't" with regard to the expediency with which one reports their rape. Jurors may find suspicion in either a speedy or delayed report, but in both cases this distrust stems from improper, yet ingrained, assumptions and pseudo-psychology based on myths as to what the "proper" and believable manner is in which a rape victim should act.

In an attempt to rectify this issue, Section 6 of the Abusive Behaviour and Sexual Harm (Scotland) Act 2016 introduced a statutory requirement in sexual offence cases for the judge, when charging the jury, to advise them that there may be perfectly valid reasons for the victim of a sexual offence to not tell others about it, or delay in doing so; and perfectly valid reasons for the victim to not utilise physical force against the performer of the illegal act. In doing so, the judge has a statutory duty to direct the jurors that a delay in

[54] ibid 208.
[55] ibid.
[56] ibid 209.
[57] ibid 210.

reporting, or the absence of physical resistance, do not in essence negate the veracity of the alleged offence.[58] While this may be an effective idea in theory, attempting to dissuade years' worth of deeply engrained attitudes and beliefs is no easy task. This is particularly given the propensity of jurors to not take on board judicial directions, which will be discussed in greater detail anon. As such, while the Act is a noble endeavour by legislators on a theoretical level, its effectiveness in practice may be limited.

The prevalence and consistent effect of these rape myths on jurors is particularly problematic in Scotland. As aforementioned, the offence of rape remains under the exclusive remit of the High Court and is always tried under solemn proceedings. This means that all cases of rape and sexual assault in Scotland are tried before a jury, with no option for the case to be tried without them. The evidence of mock jury studies has reflected a consistent reluctance among jurors to believe individuals who claim they have been raped, often regardless of evidence provided in support of their claims. Therefore, it seems that the use of juries in rape trials has damaging consequences on the application of proper justice for those who have suffered one of the most heinous crimes known to man. It is of no surprise that politicians,[59] academics,[60] and journalists[61] have called for juries to be abolished in cases of sexual assault. The attitudes held by wider society, which leak into juries, are often simply too prejudiced to allow for a fair trial.

Prejudice from Publicity

As well as being problematic in cases concerning rape myths and race, jurors' impartiality has also been shown be badly affected by pretrial publicity (PTP) in individual cases. Scotland, along with the rest of the free world, has overseen a raging battle between the freedom of the press and the right of an accused to a fair trial, both of which are

[58] Abusive Behaviour and Sexual Harm (Scotland) Act 2016, s 6.
[59] HC Deb 21 November 2018, vol 649, cols 347-348.
[60] Callander (n 10).
[61] Julie Bindel, 'Juries Have No Place In Rape Trials. They Simply Can't Be Trusted' *The Guardian* (London, 21 November 2018) <www.theguardian.com/commentisfree/2018/nov/21/juries-rape-trials-myths-justice> accessed 2 March 2020.

protected under the ECHR, Articles 10 and 6. It has been held by the ECtHR that the media is allowed freely to inform the public of matters relevant to ongoing proceedings, insofar as the authority and impartiality of the court is not undermined.[62] Nonetheless, the domestic courts have highlighted the threat which PTP poses towards juries' obligation to determine cases purely on the facts presented to them in court. The leading case on this matter in Scotland is that of *Stuurman v HMA*.[63] Jan Stuurman, amongst others, was accused of drug related offences. Immediately following their arrest, publications were aired describing the arrest, suggesting that the accused had committed other criminal acts in the Netherlands, and that those arrested were an international gang with criminal connections.[64] It was held by the court that a case could be prevented from proceeding if it was found that requiring the accused to face a jury trial amidst prejudicial publicity would amount to oppression.[65]

The PTP in *Stuurman* was held not to amount to oppression,[66] despite the fact that a broadcasting company involved in the publications was found guilty of gross contempt of court.[67] Nevertheless, jury research has indicated that PTP can indeed have a biasing effect on jurors, present even in the face of consistent reminders not to consider PTP. In their study, Ruva and LeVasseur conducted an experiment involving two sets of fourteen juries, with one set exposed to fictional PTP relevant to the case.[68] The exposed set were strictly instructed not to give any consideration to the PTP in their deliberations or their final verdict.[69] However, the results reflected that, at some point in deliberations, all fourteen juries in the exposed condition mentioned at least one fact of PTP, with four explicitly mentioning that the PTP was influencing their verdict.[70]

---

[62] *Sunday Times (No 1) v United Kingdom* App no 6538/74 (ECtHR, 26 April 1979).
[63] 1980 JC 111.
[64] ibid.
[65] ibid.
[66] ibid.
[67] ibid.
[68] Christine L Ruva and Michelle A LeVasseur, 'Behind Closed Doors: The Effect Of Pretrial Publicity On Jury Deliberations' (2012) 18 Psychology, Crime & Law 431.
[69] ibid.
[70] ibid.

Further, and even more shockingly, 72% of the juries in the exposed set delivered a guilty verdict, in stark contrast to a mere 19% in the non-exposed condition.[71]

This startling research shows the damage that PTP can do to the effective application of justice when juries are involved. As mentioned earlier, the expectation in the Scottish courts is that jurors will listen to their judicial instructions not to allow PTP to affect their deliberations. However, the above study has shown that this rarely occurs in practice. Further, the Contempt of Court Act 1981 prevents any revelations of jury deliberations outside the jury room and ensures that any PTP considerations that do arise will remain confidential. This is particularly dangerous given the leap in guilty verdicts demonstrated in Ruva and LeVasseur's study when there is negative PTP. It is important to note that this effect may be lesser in Scotland; indeed its laws concerned with PTP have been described by a practising lawyer as 'the most restrictive in the world'.[72] Nevertheless, real world cases, such as *Stuurman,* have shown that negative PTP can exist in Scotland, and that the courts are reluctant to find in favour of the accused regarding claims of prejudicial bias arising from PTP. As such, the fact that juries are prone to PTP bias is a real danger to the impartial application of justice. This is particularly true in the social media age, where potentially prejudicial information is so easily accessible.

Summary

There are multiple issues with the criminal jury system in Scotland. The duty of the trier of fact in criminal trials is to deliberate intelligently on the facts presented to them in the course of the trial, and deliver a fair and legally sound verdict on the basis of those facts, and those facts alone. As has been demonstrated above, there are many risks which arise from the practice of laypersons being granted this power. The evidence presented through mock juror studies reveals that jurors often lack the competence to understand the

[71] ibid.

[72] Alastair Bonnington, 'The Stupidest In The World?' (*Law Society of Scotland*, 2004) <www.lawscot.org.uk/members/journal/issues/vol-49-issue-12/the-stupidest-in-the-world/> accessed 14 March 2020.

complex legal concepts and directions they are given, and their impartiality may be affected by biases arising from racial prejudice, rape myths, and media publicity surrounding the trial. Further, all of these potentially damning faults have been secluded from proper consideration time and time again by way of the extensive protection offered by the doctrine of confidentiality of jury deliberations. Clearly, the Scottish jury, in its current form, is not fit for purpose.

Potential for Jury Reform

While the case for the unfitness of the jury is strong, it is important to remember that the jury is a cornerstone of Scots criminal law, and has been for centuries.[73] Furthermore, the jury is still considered by many to be inherently valuable; a symbol of public engagement in the criminal justice system, legitimising the power of the state through the input of regular citizens.[74] This argument does hold some weight, bringing to mind the oft-quoted legal aphorism 'justice must not only be done; it must be seen to be done.'[75] It seems that juries, despite featuring in less than 1% of criminal trials in Scotland,[76] have both public support and a role to play in public validation of criminal justice. As such, for the extreme call for their full abolition to have any merit, it is necessary to discuss whether potential reform of the jury system, shy of abolition, could go far enough to solve its inherent problems.

Juror Competence

In the wake of the Post-Corroboration Safeguards Review of 2015, the government commissioned a report, composed by Chalmers and Leverick, which, *inter alia*, intended to identify methods by which juror understanding of evidence and judicial directions might be

[73] Willock (n 1).
[74] Peter Duff and Mark Findlay, 'Jury Reform: Of Myths And Moral Panics' (1997) 25 International Journal of the Sociology of Law 363.
[75] The origin of this aphorism is believed to be *R v Sussex Justices, ex parte McCarthy* [1924] 1 KB 256, 259 per Lord Hewart CJ: 'it is not merely of some importance but is of fundamental importance that justice should not only be done, but should manifestly and undoubtedly be seen to be done.'
[76] Duff and Findlay (n 74).

enhanced.[77] In their review, Chalmers and Leverick set out a variety of potential reforms, intended to better facilitate juror understanding and, by extension, competence. Those opined here to have the greatest potential will be discussed.

Audio-Visual Aids

The review remarks upon the findings of Brewer and others, in their research concerning the effect of visual aids on Australian jurors' understanding of legal concepts and directions. The study involved mock jurors, composed of a combination of law students and novices with no legal expertise who were given a case summary, followed by ten minutes of legal direction on the law of self-defence. This was delivered by a trial judge, and was presented in either purely auditory form, or accompanied by computer animations which highlighted key words and involved human figures which illustrated key concepts. The comprehension of the jurors was then measured using a questionnaire, with the results showing that, while the visual aids had little effect on the comprehension levels of the law students, it significantly improved the comprehension of the novices, raising it to a similar level as that of the students.[78]

Concerns were expressed over the realism of the study,[79] but nonetheless, it may, to an extent, be relied upon to highlight the potential positive effects that more visual methods of conveying legal concepts could provide in jury trials. Given that the vast majority of juries, by design, do not contain individuals with specific legal expertise, providing jurors with animations explaining the key legal concepts of the case at hand could go some way to improving their understanding of those legal concepts. This would result in better juror comprehension, and better application of the law as it is intended. Indeed, the introduction of audio-visual aids is unlikely to met with much resistance since audio-visual technologies are already

[77] James Chalmers and Fiona Leverick, *Methods Of Conveying Information To Jurors: An Evidence Review* (Scottish Government 2018).
[78] Neil Brewer, Sophie Harvey and Carolyn Semmler, 'Improving Comprehension Of Jury Instructions With Audio-Visual Presentation' (2004) 18 Applied Cognitive Psychology 765.
[79] Chalmers and Leverick (n 77).

involved in the Scottish courts to a great degree, such as the use of live television links for vulnerable witnesses.[80]

However, there are multiple drawbacks with this potential reform. Such animations must be developed for the courts, which would likely involve considerable time and expense in order to produce effectively explanatory animations for the vast swathes of legal concepts that arise in criminal trials. Further, these animations must be satisfactorily neutral in their presentation of the concepts, without the risk of further prejudicing jurors, a fear expressed by Brewer and others themselves.[81] Finally, the actual efficacy of these animations may be called into question, because while they were effective at raising the comprehension of the novice jurors, it did so only to the extent of putting them on a level footing of law students. While such students may indeed have a good degree of understanding of legal concepts, this understanding is likely pale relative to that of a professional judge, who has had far greater experience applying these concepts in practice. It seems that even with the additional support of visual aids, the comprehension of legal concepts by laypersons will be limited, particularly when compared to that of a professional judge. The proper application of these concepts by a jury in reaching a verdict will be limited by extension.

Pre-Instruction

The potential reform most strongly supported by Chalmers and Leverick was to provide the jury with instructions detailing the specific legal issues at hand prior to the presentation of evidence.[82] Citing mock jury research by Smith, it was noted that, when jurors were presented with both pre and post-trial instructions, they exhibited significantly better comprehension levels than those of jurors who had been merely post-instructed.[83] This potential reform appears to have some merit, as the results from its use in research are positive, and the implementation should not prove difficult. It would

[80] Vulnerable Witnesses (Scotland) Act 2004, s 20.
[81] Brewer, Harvey and Semmler (n 78).
[82] Chalmers and Leverick (n 77).
[83] Vicki L Smith, 'Impact Of Pretrial Instruction On Jurors' Information Processing And Decision Making.' (1991) 76 Journal of Applied Psychology 220.

be of little hardship to the judge to simply deliver instructions twice to the jurors; first, before trial to allow them to recognise the relevant legal issues; and, second, after trial for reiteration and to make allowance for any unexpected issues which may have arisen. There is a logical risk which arises in that jurors exposed to pre-trial instruction may reach a verdict within their own head before the conclusion of the trial. However, Smith's research actually found the opposite was true, with 25% of the subjects in her study who were provided with pre-instructions deferring their verdict at the midway point, as opposed to 0% of the subjects who had not.[84]

While pre-instructions do appear to allow for better comprehension of legal concepts by jurors, with no discernible drawback, it is important to note that, again, this comprehension is limited. Reflecting on the results of her study, Smith specifically notes that '[o]n all three measures, performance was quite poor, leaving considerable room for improvement in jurors' comprehension.'[85] The value of pre-instructions in increasing juror comprehension is not to be dismissed. However, their effect on improving comprehension amongst lay jurors is undeniably limited. This lack of comprehension is demonstrably both dangerous to justice and seemingly very difficult to overcome.

Removing Jury Secrecy

It is opined above that the most problematic issue with the current Scottish jury is the doctrine of confidentiality. However, this issue also appears the most easily rectifiable. As aforementioned, the Contempt of Court Act 1981 makes it a contempt of court to obtain or disclose particulars expressed in jury deliberations.[86] The compounding risks that this creates with regard the unaccountability of jurors, even in the case of jury impropriety, have been made clear. As such, a potential solution is to simply abolish Section 8 of the Act and allow for the open dissection of juror deliberations. Such openness would provide multiple benefits to the jury system, allowing for greater clarity and finality in verdicts, since the reasoning behind the jury's decision

[84] ibid 225.
[85] ibid 224.
[86] Contempt of Court Act 1981, s 8.

could be publicly presented. Further, the propensity of jurors to extrajudicial biases and low levels of comprehension would no longer be shrouded in secrecy but exposed, allowing for the correction of inappropriately influenced verdicts.

Indeed, even bolder moves have warranted some merit in academic research, such as the video recording of jury deliberations. Bagley recounts a landmark event in 1986 where jury deliberations in a criminal trial were knowingly recorded and broadcast on the American television channel PBS.[87] In contrast to potential fears that the cameras would limit the jurors in their frank and open discussions, the jurors appeared to have '…forgotten about the presence of the cameras' and 'did not appear apprehensive with respect to discussing any of the issues during deliberations.'[88] Indeed, rather shockingly, the jury, in full knowledge of the recording cameras, verbalised an intention to ignore the legal evidence and acquit the defendant on the basis that, in their opinion, to do otherwise would be unjust.[89] Such a process is known as "jury nullification", and while this issue is beyond the scope or remit of this article, it is of relevance to note that jurors may reach verdicts through non-legal reasoning, even when knowingly recorded. Such revelations would be unlikely to appear given the current rule of jury confidentiality, despite the fact that they represent a conscious move away from the oath taken by jurors to give a true verdict according to the evidence. Regardless of one's views towards nullification, it is in the interests of justice that such a rebuke of the oath should be made public.

In theory, the abolition of the Section 8 protections, accompanied by the recording of jury deliberations, would allow for better understanding of verdicts delivered by juries. There would be greater faith in the justice delivered by them and better safeguards against biased decisions and juror impropriety and extrajudicial considerations. Even if jury deliberations are not revealed in their entirety, an alternate reform would be to provide jurors with a list of yes/no questions that the jurors must answer in their deliberations.

---

[87] William R Bagley, 'Jury Room Secrecy: Has the Time Come to Unlock the Door' (1999) 32 Suffolk U L Rev 481.
[88] ibid 486.
[89] ibid.

This is standard jury practice in some jurisdictions, such as Belgium and Spain.[90] Such questions would be intended to enhance understanding of the reasonings of jury verdicts and to focus the jurors' thinking, 'spelling out' the issues for their attention and allowing them to better consider them.

However, the removal of jury confidentiality, whilst desirable in theory, is likely to be met with difficulties in practice. Despite its flaws, there are compelling reasons for retaining the principle. The most convincing reason, it is opined here, is the freedom for frank discussion. While the televising of jury deliberations in the aforementioned American case was shown to have not affected the confidence of the jurors to discuss all areas of the case, it is important to note that this was not a particularly controversial case. It concerned the unlawful possession of a firearm and as such was not particularly heinous; it also did not involve well-known individuals.[91] Jury confidentiality is intended to protect jurors from public backlash in controversial cases, such as where jurors wish to acquit an unpopular accused or convict a popular one. The uniform anonymity of jury deliberations shields individual jurors from suffering retribution for their decision in their personal life. This is particularly important because jurors do not request or necessarily desire to serve on the jury, but merely have the misfortune to be randomly selected. This is in contrast to professional judges who have made a conscious career choice and, by extension, have accepted the potential negative effects unpopular verdicts may impart.

Even the less extreme reform considered here of providing the jurors with yes/no questions, while currently used in some foreign jurisdictions, is unappealing as a potential approach in Scotland. The nature of the civil law tradition Spain and Belgium means that juries are removed from the more adversarial and oral environment of juries in Scotland. It is therefore difficult to form an effective comparison. The lack of a meaningful comparator aside, introducing such a process in the Scottish jury system would open up a swathe of additional

[90] Stephen C Thaman, 'Should Criminal Juries Give Reasons for Their Verdicts?: The Spanish Experience and the Implications of the European Court of Human Rights Decision in Taxquet v. Belgium' (2011) 86 Chi Kent L Rev 620.
[91] Bagley (n 87).

problems. The issue of creating the question sheet posed to juries runs the risk of the wording of the questions being unfairly prejudicial to either the accused or alleged victim. Issues remain even if the questions are neutrally worded. Requiring jurors to deliberate through the lens of yes/no questions may result in jurors reaching their verdict by simple checklist, or forcing difficult considerations into a Procrustean Bed of sorts, whereby nuanced and complicated considerations are reduced to a simple yes or no.

In 2021, the Lord Justice Clerk's Review Group considered the above reform in relation to sexual offence cases. The Review Group wholeheartedly rejected the proposal. It noted the confusion which may arise from conflicting answers between jurors; the greater length of time which would arise from such formalisation of jury deliberations; and the risk of an increased number of appeals.[92]

Jury confidentiality is seemingly untouchable insofar as the jury system remains in place. It is necessary for the protection of jurors who, through no choice or fault of their own, have been placed in a position where they face potential public backlash. On the other hand, jury confidentiality inhibits effective and reliable justice. This is a Catch-22 situation where reform is both necessary and impossible. This is a strong argument in favour of the abolition of the Scottish jury system.

Targeting Jury Prejudice

As set out in chapter one, the reliance on judicial directions and a hopeful attitude is failing to curb jurors' propensity to allow prejudicial biases to influence their deliberations. One potential reform which may improve jury impartiality is a *voir dire* process, which involves vetting potential jurors prior to trial. In Scotland, there is no such process, though its use is extensive in the United States.[93] There, potential jurors are routinely and thoroughly vetted by both the prosecution and defence, being asked questions concerning their

92 Lady Dorrian, *Improving the Management of Sexual Offence Cases: Final Report from the Lord Justice Clark's Review Group* (Scottish Courts and Tribunals Service 2021).
93 Duff (n 2) 180.

personal attitudes and practices.[94] The theoretical justification for this is to weed out potentially prejudicial biases, on either side, which may adversely affect juror impartiality. Replicating such a process in Scotland could be potentially useful in removing problematic jurors who may hold prejudicial biases arising from their personal views, belief of rape myths, or exposure to PTP. Such removals would go a great way to allowing for greater impartiality from juries, and thus greater justice.

However, the American use of *voir dire* has attracted criticism in practice. Rather than remove biased jurors, it has been accused of planting them. That is to say, defence attorneys' questioning of jurors in the *voir dire* process is not intended to create an impartial jury, but rather to identify jurors partial to their client and ensure their presence on the jury.[95] This criticism is convincing. A lawyer acts in his client's best interests. When presented with an option to legally stack the odds in his client's favour, it should be of no surprise that he should take it. *Voir dire*, rather than lessening the impact of juror prejudice, instead validates it, inviting it in through an open door. The Scottish courts have been critical of suggestions to introduce *voir dire* to Scotland, citing the prevalence of its abuse. Returning to *McCadden*, Lord Wheatley remarked:

> There may never be a process which eliminates the possibility of personal prejudices existing among jurors, the nearest practical one… being possibly the "vetting" of jurors, a system against which the law of Scotland has steadfastly closed the doors. Evidence of how it is used and abused in countries in which it is operated only tends to confirm the wisdom of that decision.[96]

It can safely be assumed that *voir dire* will never feature in Scots criminal law. Given the risk it poses to worsen injustice, this may be a

---

[94] Albert W Alschuler, 'The Supreme Court And The Jury: Voir Dire, Peremptory Challenges, And The Review Of Jury Verdicts' (1989) 56 University of Chicago L Rev 153, 158.
[95] Brittany L Deitch, 'The Unconstitutionality of Criminal Jury Selection' (2018) 26 William and Mary Bill of Rights Journal 1059, 1069.
[96] *McCadden v HM Advocate* 1985 JC 98.

positive. *Voir dire* is not the solution to the issue of prejudicial biases among jurors. For as long as this issue festers, the role of juries will continue to be called into question.

Summary

This chapter has considered potential reforms to the problematic jury in Scotland short of abolition. However, these reforms would be either too limited in their efficacy or give rise to troublesome second-order effects. The ability of laypersons to grasp complex legal concepts, while improvable, nevertheless falls short. Abolishing the doctrine of jury confidentiality, while improving accountability and impartiality, carries with it an unacceptable level of risk for individual jurors. A vetting process, while theoretically removing problematically prejudiced jurors, in practice leads to abuse. Satisfactory reform of the jury is impossible. Given the scale of the issues with the current system, the only viable solution is abolition.

Recommendations Following Abolition

Chapters one and two have revealed that the abolition of juries is both desirable and necessary. This chapter will consider potential alternatives to the jury in the Scottish criminal trial. While it has already been noted that jury trials only constitute a small minority of cases, these few cases might well be considered the most important cases, because they involve the most serious offences. As such, the stakes for wrongful conviction or acquittal is highest in these cases. The effectiveness of any replacement is paramount.

UK Trials Without Juries

One need not look too far to find occasional situations in which jury trials in the UK can be circumvented by legislation. Sections 43 and 44 of the Criminal Justice Act 2003, which have force across the UK, make allowance for the prosecution to apply for the trial to be conducted without a jury in lengthy or complex fraud cases, and cases in which there is a real and present danger of jury tampering.[97] If successful, such an application results in the trial being heard by a judge alone.

[97] Criminal Justice Act 2003, ss 43-44.

While Section 43 is not yet in force, the section demonstrates an understanding on the part of the legislature of the risks that issues with juries' comprehension pose when cases are particularly complex. The solution, enshrined in statute, is to dispense with them and have a professional judge in their stead.

This was also the implemented solution for multiple cases in Northern Ireland during the Troubles. Juries were considered a risk to justice given the likelihood of jurors being influenced by fear or extreme partisanship due to the situation in Northern Ireland at the time. This led to the establishment of so-called 'Diplock Courts'.[98] Therein, trials involving certain scheduled offences, generally related to terrorism, were automatically to be conducted by a judge alone, without a jury. The trial was deemed to have the same authority as if conducted with a jury.[99] Following the end of the Troubles, the notion of scheduled offences automatically leading to a trial without a jury was abolished by the Justice and Security (Northern Ireland) Act 2007. However, the 2007 Act still allows the Director of Public Prosecutions for Northern Ireland to issue a certificate for a trial to continue without a jury if the defendant or offence is considered to be linked to a terrorist organisation operating within Northern Ireland, and this connection is believed to create a risk regarding the proper administration of justice.[100] In this case, again, it seems that the solution concerning the prejudicial bias problem in jurors was to replace them with a professional judge.

More recently, the Scottish Government gave real consideration to the introduction of emergency legislation related to the ongoing coronavirus pandemic which would have seen criminal trials proceed without juries, presided over only by a professional judge.[101] While these plans were dropped in the face of public

[98] John Jackson and Sean Doran, 'Diplock And The Presumption Against Jury Trial: A Critique' [1992] Crim LR 755.
[99] Northern Ireland (Emergency Provisions) Act, 1973, s 2.
[100] Justice and Security (Northern Ireland) Act 2007, s 1.
[101] David Bol, 'Coronavirus In Scotland: No-Jury Trial Proposals Labelled "Draconian"' *The Herald* (Glasgow, 31 March 2020) <www.heraldscotland.com/news/18347994.coronavirus-scotland-no-jury-trials-proposals-labelled-draconian/> accessed 1 April 2020.

backlash,[102] they still demonstrate a legislative propensity towards the use of professional judges as the potential replacement of juries in Scotland in certain circumstances.

The Criminal Justice Act 2003 and the Justice and Security (Northern Ireland) Act 2007 show that the UK government appears occasionally ready to dispense with the use of juries in select criminal trials. The alternative on all occasions has been to fill the gap left behind by the jury with a professional judge. However, this situation only applies to select and often extreme cases. Were the jury to be abolished completely in all cases, as this article proposes, then the use of professional judges would be extended across the board to all cases, extreme or not. In determining whether this would be a satisfactory move, it is useful to examine the effect this has had in a jurisdiction that has already adopted such measures.

Comparisons with South Africa

Despite their crucial differences, the Scottish jury has been described as 'more closely related to the English jury than any other modern system of lay participation in the criminal process.'[103] As such, the leap may be considered not so great between the two systems. This allows useful comparisons to be drawn between the abolition of the long-standing English style jury system of South Africa and the potential abolition of juries in Scotland. Scotland and South Africa both have mixed legal systems, drawing on common law and civil law traditions. They enjoy a shared history and social culture, intertwined by the Empire and presently through the Commonwealth.

The history of the jury in South Africa is a complicated one. It involves multiple European colonisers, predating the British hegemony, and their respective legal systems. The originally Dutch colonies of Natal and the Cape of Good Hope were governed by Roman-Dutch law with no institution comparable to a jury.[104]

---

[102] BBC News, 'Coronavirus: MSPs pass emergency powers bill' (*BBC News*, 1 April 2020) <www.bbc.co.uk/news/uk-scotland-scotland-politics-52111412> accessed 1 April 2020.

[103] Christopher Gane, 'The Scottish Jury' (2001) 72 Revue internationale de droit pénal 259.

[104] Richard Vogler, 'The International Development Of The Jury: The Role Of The British Empire' (2001) 72 Revue internationale de droit pénal 525.

Conversely, Boer settlers in the Transvaal and the Orange Free State colonies did have lay involvement in criminal justice. This was not uniform, however. The Transvaal utilised twelve-person juries and lay judges, while the Orange Free State utilised a jury of six or nine.[105] Later, British rule in South Africa resulted in the imposition of English law and English-style juries across the region. As a result, by the time the four colonies were united into the Union of South Africa[106] in 1910 by the South Africa Act 1909,[107] the empanelling of nine-person juries for serious offences was ubiquitous throughout the dominion, albeit with different rules regarding jury unanimity and eligibility.[108] Finally, the Criminal Procedure & Evidence Act No 31 of 1917 completely standardised the South African jury, with eligibility limited to parliamentary voters, and majority verdicts of seven to two being set as the minimum.[109]

However, no sooner had the structure of the South African jury been settled than it began to experience a piecemeal dismantling. Even within the 1917 Act, there were provisions allowing the accused to insist on trial without a jury.[110] Further, a 1935 amendment allowed the Minister for Justice to order trials without a jury, an allowance which was utilised liberally.[111] A continuing lack of confidence in the jury system led to an amendment to the criminal code in 1954 which stipulated that trial without a jury would be the standard, unless the accused specifically requested one.[112] The last nail in the coffin of the jury in South Africa was the Abolition of Juries Act 34 1969, which officially repealed all laws relating to the use of juries in criminal proceedings.[113] Practically, however, the jury in South Africa had fallen into disuse long before its official abolition, replaced instead

[105] ibid 542.
[106] Historical predecessor to the present day nation of South Africa.
[107] South Africa Act 1909.
[108] Vogler (n 104), 542.
[109] Criminal Procedure & Evidence Act 31 1917.
[110] ibid s 216.
[111] South African Law Commission, *Simplification Of Criminal Procedure (Access To The Criminal Justice System)* (South African Law Commission 1997) 16-17.
[112] ibid 17.
[113] Abolition of Juries Act 34 1969.

with a professional judge, occasionally accompanied by up to two lay assessors. This remains the case today.[114]

The lack faith in the jury system in South Africa, and its eventual abolition, have been attributed to the South African jury's uniquely racial history. Whilst non-whites did appear regularly on juries in the very early days of the Cape colony,[115] South African juries were generally, and later exclusively, white. Regardless of obvious issues of justice with completely homogeneous juries in any jurisdiction, this was of particular issue in South Africa with its black majority and ubiquitous racial discrimination. Commenting on juries in nineteenth century Natal, Spiller went so far as to remark that 'the use of the jury system was the most crippling handicap faced by Blacks in a system that generally disadvantaged them.'[116] The racial injustice embedded in the jury system did not improve under the later policy of Apartheid. Unsurprisingly, therefore, the jury system in South Africa was abolished over five decades ago. The South African Law Commission did consider the possibility of reintroducing juries in South Africa in 1997, citing the improvement in South Africa's race relations in recent history.[117] However, the discussion was not particularly positive, with issues arising concerning increased costs, delays in justice, juror competence, jury confidentiality, prejudice, and a lack of popular desire for change.[118] It seems that South Africa is content with its non-jury system, with trials heard only by a judge, potentially accompanied by lay assessors. Particularly noteworthy is that South African judges are not only permitted to reveal their reasoning for reaching their verdicts, but are statutorily obligated to do so by virtue of the South African Constitution.[119] This is intended to promote transparency and foster trust in the criminal justice system. As discussed previously, extending this obligation, and its

[114] South African Law Commission (n 111).
[115] Milton Seligson, 'Lay Participation In South Africa From Apartheid To Majority Rule' (2001) 72 Revue internationale de droit pénal 273, 274.
[116] P R Spiller, 'The Jury System in Early Natal (1846-1874)' (1987) 8 J Legal Hist 129, 143.
[117] South African Law Commission (n 111).
[118] ibid 18-19.
[119] Hilke Thiedemann, 'The Lawful Judge – A comparative survey on the allocation of cases to judges in South Africa and Germany' (2003) 36 Verfassung in Recht und Übersee 257.

associated benefits, to the Scottish jury system would be an incredibly difficult task.

When attempting to apply the South African model to Scotland, some difficulties arise. Juries have been a stalwart of Scots law for centuries; whereas in South Africa they could be seen as a recent and foreign imposition. Further, juries have received public support in Scotland, whereas at no point in South African history were juries particularly popular. Finally, the racial element of South Africa's history and its extreme negative effect on the jury system is unique to South Africa. Difficulties may be had attempting to extrapolate the South African experience elsewhere. Nonetheless, South Africa's successful abolition of juries from its criminal justice system bodes well for similar reform in Scotland.

Although Scotland has never suffered government-enforced racial discrimination comparable to South African Apartheid, its history is still plagued with religious sectarianism.[120] Indeed, Scottish religious strife has been shown to have potentially leaked through into juror impropriety.[121] If South Africa's deeply ingrained racial problems were improved by the replacement of juries with judges, as seems to be the prevailing view, then such a move in Scotland could be expected to be particularly effective. Professional judges can be regarded as an effective solution to the unsatisfactory jury system. They are already subject to a procedural safeguards against improper verdicts through their professional and public accountability. There is much more at stake for a professional judge who may risk losing his professional standing or even career if he was revealed to be prejudicial. This is not to say that professional judges are sure to be unbiased beyond reproach, but the risk of improper factors affecting their deliberations appears far lower than those of juries. As such, the adoption of South Africa's judge-only system, potentially accompanied by lay assessors to offset complaints of an absence of community involvement in the Scottish criminal justice system, appears to be a workable and desirable replacement which has been shown to work in a comparative legal system.

---

[120] Glasgow City Council, *Sectarianism In Glasgow: Final Report* (NFO Social Research 2003).
[121] *McCadden v HM Advocate* 1985 JC 98.

Summary

The jury system in Scotland cannot be adequately reformed. It must be replaced entirely. In certain circumstances, the legislation has been enacted in the UK which has replaced jury trials with judge-only trials. This is clearly Parliament's instinctual preferred alternative to jury trials. The implementation of judge-only system is widely regarded as having improved criminal justice in South Africa. Due to the similar legal systems and shared history, useful comparisons can be drawn between South Africa and Scotland. Scotland should follow South Africa's lead and replace the jury with a judge.

Conclusion

The use of juries in Scottish criminal trials dates back centuries. It has been a popular and enduring feature of Scots criminal law almost since its inception. It is seen as a symbol of democracy and balance against the authoritative power of the state. However, this popular image does not reflect the reality of the shortcomings of jury trials. Laypersons, unqualified in the complexities of the law, have consistently demonstrated inadequate legal comprehension and high prejudicial bias. Further, the doctrine of jury confidentiality leads to a lack of accountability. This doctrine is preserved in statute and strengthened by case law. Jury confidentiality heightens the dangers inherent in jury trials.

Options considered for reform of the jury, while noble in their aims, have been shown to fall short of effective change. The only adequate solution is the cessation of the involvement of juries in criminal trials. They should be replaced as triers of fact by professional judges, potentially supported by lay assessors. Professional judges have experience and training to properly understand and apply the law; are better suited to appreciate the risks which prejudicial biases pose to justice; and are much more accountable for their decisions. A system whereby professional judges replace juries has precedent in the UK under the Criminal Justice Act 2003 and the Justice and Security (Northern Ireland) Act 2007. The idea has also been shown to work well in South Africa, which is a useful comparator jurisdiction to Scotland. Even in Scotland itself, legislation was proposed in

response to the coronavirus pandemic which would have seen judge-only trials implemented as an emergency measure. Proponents of the jury system appear to have defeated these plans, defending it on its symbolic level. However, symbolism alone is not enough to discount the risks to the competence, fairness, and propriety which the Scottish criminal justice system demands. Lord Devlin referred to juries as 'the lamp that shows that freedom lives.' It appears, upon further reflection, that lamplight had blinded him to the danger.

Cyber-attacks as an Emerging Use of Force under International Law

ENENU ONYIKWU OKWORI*

Abstract

The digital age presents threats as well as opportunities. It is becoming increasingly important to understand how computer systems can be used in harmful cyber operations. Cyber-attacks can be devastating for individuals, organisations and even states. Recently, arguments have arisen as to the extent to which existing international law, particularly international law on the use of force, applies to this evolving field of cyberspace operations. This article argues that in certain circumstances, a cyber-attack can satisfy the criteria to be considered a use of force or even an armed attack under the UN Charter.

Keywords: cyber-attacks, use of force, international law

Introduction

In recent years, development has largely been marked by an increased dependence on technology. The need for interconnectedness and a digitalised global economy has triggered more reliance on the internet, which is:

> a global computer network providing a variety of information and communication facilities, consisting of interconnected networks using standardized communication protocols.[1]

A 2021 study shows that about 4.66 billion of the world's population make use of the internet, which is about 59% of the total

* LLB (Nigeria), BL (Nigeria), LLM (UK). The author is a final year PhD candidate at the University of Aberdeen. Her research interests are primarily international law, including the use of force, cyberspace and technology, and international human rights law.

[1] *The Oxford Dictionary of Phrase and Fable* (2nd edn, OUP 2006).

world population. Out of this number of people, over 92% of them access the internet using mobile devices.[2]

This global digital world is now being referred to as the 'internet of things'. This describes everything and anything connected to the internet, so much so that everyday life simple machines now 'talk' to each other. Phones, wristwatches, TVs, sound systems, mechanised controls, etc are all linked together; communicating and sharing data through internet connections.[3]

Cyber-attacks, cyber-hacks, cybercrime, cyber terrorism: these are all risks which could erupt at any time. These are the downsides of technological advancement in our present globalised world. Current trends show that these harmful cyber operations are carried out against governments, organisations, corporations and even individuals at an alarming rate and which will continue to rise.[4] These challenges notwithstanding, the perks that come with this advancement are highly advantageous. Entities which are likely to be affected by cyber disruption should do all in their power to repel these attacks by boosting their cyber security and strategies towards prevention, detection and even remediation.

Alongside these strategies, there is a widely held view that the solution to these issues lies partially in international law. This is especially because states are major targets of these attacks, which can be perpetrated by other states or non-state actors. The debate has shifted from whether international law in general applies in cyberspace operations to how existing fundamental principles or rules should actually apply. While some have advocated that entirely new rules or principles should be formulated to create a separate law of cyberspace, others have suggested persuasively that the existing principles and rules of international law should apply, albeit with

---

[2] Joseph Johnson, 'Global Digital Population as of January 2021' (*Statista,* 7 April 2021) <www.statista.com/statistics/617136/digital-population-worldwide/> accessed 18 April 2021.

[3] Matt Burgess, 'What is the Internet of Things?' (*Wired*, 16 February 2018) <www.wired.co.uk/article/internet-of-things-what-is-explained-iot> accessed 18 April 2021.

[4] *Cyber Security Report 2020* (Check Point Research 2020).

modifications where necessary.[5] One of the prevailing arguments particularly has been whether the law on the use of force can also be applied to cyber operations like any other kinetic attack. The reasoning is that if cyber-attacks are classifiable as a use of force or even an armed attack under international law, then it would be permissible for states to undertake proportionate counter measures or self-defence in response, where necessary. However, the modalities and feasibility of this may not be as straight forward as it seems since cyberspace operations are largely 'unseen'. This is why some experts have advanced the view that it would be practically impossible to equate a cyber-attack to a kinetic one for the purposes of determining whether a use of force has occurred or not.

This article will briefly discuss the concept of cyber-attacks and then analyse how cyber-attacks ideally fit within the interpretation accorded to the use of force under the UN Charter. The law on the use of force is quite broad, hence this article will only consider aspects that relate to the discussion at hand: the conceptualisation of the use of force as armed force and whether by application, cyber-attacks can be termed as such. The threat of force, counter measures and the right to self-defence will also be briefly considered.

The Concept of Cyber-attacks

Debates around cyber-attacks have only gained momentum due to the recent surge of cyber-attacks in diverse forms. However, that is not to say that cyber-attacks are contemporary occurrences that have only recently been discovered. Cyber-attacks have evolved from early malware intrusions like viruses in form of the notorious 'Trojan Horse' that enabled hackers access private computers and corrupt data stored on them.[6] Now, cyber-attacks range from the minor disruptions that affect private computers and companies that often are not aware of the attacks until they have been victimised, to major attacks affecting major forums with more drastic effects, like the 2010

[5] *Trends in international law for cyberspace* (NATO CCDCOE 2019);
Michael Wood, 'International Law and the Use of Force: What Happens in Practice?' (2013) 53 Indian J of Intl L 345, 354-355.

[6] Titiriga Remus, 'Cyber-Attacks and International Law of Armed Conflicts: A *Jus ad Bellum* Perspective (2013) 8 J of Intl Com L & Tech 179.

spear-phishing attack on Google.[7] This attack affected over 75,000 systems across 2500 companies and was even declared the most sophisticated at the time.[8]

Many attempts have been made at defining a cyber-attack, although the question remains unsettled. Two creditable definitions are those provided by the UK and the US respectively. The UK's National Cyber Security Centre (NCSC) has described cyber-attacks as 'malicious attempts to damage, disrupt or gain unauthorised access to computer systems, networks or devices, via cyber means'.[9] This definition is succinct and presents the path through which many cyber-attacks are launched, but it fails to consider the effects of attacks that do not directly impact computer systems or networks. The definition provided by the US Department of Defence presents a broader perspective and takes into account the extensive impacts a cyber-attack can have. It has defined cyber-attacks as

> actions taken in cyberspace that create noticeable denial effects (i.e., degradation, disruption, or destruction) in cyberspace or manipulation that leads to denial that appears in a physical domain and is considered a form of fires.[10]

There are different types of cyber-attacks. Access attacks, for example, entail those cyber-attacks by unauthorised hackers to hack into devices. Vulnerabilities within a device or its services are exploited to gain unauthorised access to sensitive information. Examples include queries related to internet information. This attack can engage domain name server (DNS) queries to identity the owner of a domain including the internet protocol (IP) addresses attached to it. Phishing emails are another form of access attacks. Hackers send false emails to random unsuspecting IT users under the pretence of being legitimate businesses or financial institutions to lure them into

---

[7] Jack Schofield, 'Google, Yahoo, Adobe and Who?' *The Guardian* (London, 14 January 2010) www.theguardian.com/technology/2010/jan/14/google-yahoo-china-cyber-attack> accessed 27 August 2018.
[8] Eric T Jensen, 'Cyber Warfare and Precautions against the Effects of Attacks' (2010) 88 TLR 1533, 1536.
[9] 'NCSC glossary' (*NCSC*, 23 November 2016) <www.ncsc.gov.uk/information/ncsc-glossary> accessed 8 December 2021.
[10] *Joint Publication 3-12: Cyberspace Operations* (Joint Chiefs of Staff 2018) GL-4.

clicking links or imputing their personal details on unsecured sites. This allows the attacker to steal the IT users' identities or infiltrate their financial accounts.[11] Where the hackers use personal information at their disposal to make the dodgy email appear more realistic, it is known as spear phishing.[12] Spear phishing using phone calls is known as vishing. This was employed in a recent hacking undertaken against Twitter. The hackers had allegedly posed as Twitter staff claiming that they had been locked out of their work systems and requested password reminders to enable access. With this, they were able to gain access into several celebrity twitter accounts.[13]

Denial-of-Service (DoS) attacks are also another form of cyber-attack. They target computer systems, making them run slow or even crash, rendering them deliberately unusable. Information stored on such systems could also be deleted or corrupted.[14] In extreme cases, the attacks could be effected by using a 'mass of infected computers',[15] known as botnets. The botnets attack a system, sending several requests for information until the system starts to run slow or even crash, essentially denying service to users.[16] This is termed a Distributed Denial-of-Service (DDoS) attack and they are very common.

Cybercrimes are mostly dealt with by domestic criminal justice systems. 'Crimes' are committed through the use of IT; usually for financial gain, obtaining information illegally or other criminal purposes. The Council of Europe's Convention on Cybercrime[17] represents, to a large extent, the international perspective on

---

[11] 'Phishing Attacks: Defending Your Organisation' (*NCSC*, 5 February 2018) <www.ncsc.gov.uk/guidance/phishing> accessed 8 July 2019.
[12] ibid.
[13] 'Twitter hack: Staff tricked by phone spear-phishing scam' (*BBC News*, 31 July 2020) <www.bbc.co.uk/news/technology-53607374> accessed 31 July 2020.
[14] M Uma and G Padmavathi, 'A survey on Various Cyber Attacks and their Classification' (2013) 15 Intl J of Network Security 390, 393.
[15] Arie J Schaap, 'Cyber Warfare Operations: Development and Use under International Law' (2009) 64 AFL Rev 121, 134.
[16] Russell Buchan, 'Cyber Attacks: Unlawful Uses of Force or Prohibited Interventions?' (2012) 17 J of Conflict & Security L 211, 218;
Common cyber attacks: reducing the impact (NCSC 2016).
[17] Convention on Cybercrime (adopted 23 November 2001, entered into force 1 July 2004) ETS 185 (Budapest Convention).

cybercrimes. It criminalises any infringement of computer systems, copyright or related rights, computer-related fraud and forgery, child pornography and aiding and abetting the commission of any of these crimes.[18]

Cyber-attacks could either be malicious large scale or non-malicious small-scale attacks.[19] Non-malicious small-scale cyber-attacks emanate from individuals or small groups or could be as a result of accidental or operational mishaps and are often low intensity, with minimal effects.[20] Also, the cost of damage control is usually not high. On the other hand, malicious large-scale cyber-attacks are calculated to cause wide-spread harm and could also be carried out for personal gain. Such an attack could make use of several computer systems to infiltrate vulnerable systems leading to loss of data and inaccessibility.[21] Such attacks may be inexpensive to execute but at the same time could be difficult to detect and rectify when there is serious damage.[22]

Cyber-attacks as a Use of Armed Force

Article 2(4) of the UN Charter[23] provides that:

> all members shall refrain in their international relations from the threat or use of force against the territorial integrity or political independence of any state, or in any other manner inconsistent with the Purposes of the United Nations.

Although the UN Charter has not ascribed any specific definition to 'force', the most widely accepted interpretation is that article 2(4) was

---

[18] ibid arts 2-11.
[19] Uma and Padmavathi (n 14) 394.
[20] ibid
[21] ibid
[22] Jensen, *Cyber Warfare* (n 8) 1536
[23] Charter of the United Nations (adopted 26 June 1945, entered into force 24 October 1945) 1 UNTS 16 (UN Charter) art 2(4). The importance of this prohibition was further reiterated: UNGA, 'Declaration on the Enhancement of the Effectiveness of the Principle of refraining from the Threat or Use of Force in International Relations' UN GAOR 42nd Session Supp No 49 UN Doc A/RES/42/22 (1988).

clearly referring to the prohibition of 'armed force' only.[24] This involves, in particular, the military force of any state; that is, its armed forces or non-state actors acting by or on its behalf.[25] This interpretation may be because the UN Charter has a strong focus on preventing future generations from experiencing the horrors of World War II.[26] Of course, World War II was fought with armies, navies and air forces.[27]

The preamble[28] of the UN Charter also mentions that the use of 'armed force' shall be limited to circumstances of 'common interest'. Articles 41 and 44 of the Charter refer to 'armed force' in dealing with measures that may be taken by the Security Council to restore order.[29] Another basis for this interpretation could be that other forms of force, such as economic coercion, were rejected at the preparatory level of the Charter. This supports a stricter and more explicit meaning of 'force' with regards to article 2(4).[30]

---

[24] Tom Ruys, 'The Meaning of "Force" and the Boundaries of the *Jus ad Bellum*: Are "Minimal" Uses of Force Excluded from UN Charter Article 2(4)?' (2014) 108 American J of Intl L 159, 163;
Yoram Dinstein, *War, Aggression and Self-Defence* (6th edn, CUP 2017) 90;
Ian Brownlie, *International Law and the Use of Force by States* (OUP 1963) 362;
Derek W Bowett, *Self-Defence in International Law* (Manchester University Press 1958) 148;
Albrecht Randelzhofer and Oliver Dorr, 'Article 2(4)' in Bruno Simma and others (eds), *The Charter of the United Nations: A Commentary*, vol 1 (3rd edn, OUP 2012) 208.

[25] In the end, the state bears responsibility for actions taken by it or on its behalf: 'Draft articles on Responsibility of States for Internationally Wrongful Acts, with Commentaries' (2001) UN Doc A/56/10, arts 4-10;
Brownlie (n 24) 361;
Oliver Dorr, 'Use of Force, Prohibition Of' in Rüdiger Wolfrum (ed) *Max Planck Encyclopedia of Public International Law (MPEPIL)* (OUP 2015) para 15;
Armed Activities on the Territory of the Congo (Democratic Republic of Congo V Uganda) (Merits) [2005] ICJ Rep 168 (*Armed Activities* Case) para 162;
Case concerning Military and Paramilitary Activities in and against Nicaragua (Nicaragua v United States of America) (Merits) [1986] ICJ Rep 14 (*Nicaragua* Case).

[26] UN Charter (n 23) Preamble.

[27] Brownlie (n 24) 362.

[28] UN Charter (n 23) Preamble 7.

[29] ibid arts 41, 44.

[30] This is the prevalent view under international law: Ruys on 'The Meaning of "Force" (n 24) 163; Randelzhofer and Dorr (n 24) 208

Such a restricted interpretation provides a more realistic approach for state practice and general international law. Apart from this view widely accepted by states, it also limits the circumstances in which states can lawfully resort to armed force. However, it is obvious that the drafters of the Charter did not envisage any new threat emerging outside the conventional interpretation of force that could pose a fundamental challenge to international peace and security.

In recent times, there have been several instances of alleged cyber-attacks, perpetrated by states and non-state actors alike, against states. In 2007, a major cyber-attack, in the form of a DDoS attack, struck the state of Estonia, causing several disruptions to major websites and an overall impairment on the operations of major governmental organisations.[31] The then Estonian President had remarked in an address to the United Nations General Assembly that:

> cyber-attacks are a clear example of contemporary asymmetrical threats to security… a threat not only to sophisticated information technological systems, but also to a community as a whole… The threats posed by cyber warfare have often been underestimated.[32]

In 2016, Russia allegedly perpetrated cyber-attacks targeting power grids in Ukraine[33] and, more recently, Wiper cyber-attacks as a

---

[31] Samuli Haataja, 'The 2007 Cyber-Attacks against Estonia and International Law on the Use of Force: an Informational Approach' (2017) 9 Law, Information & Tech 159, 160-161;
Herbert Lin, 'Cyber Conflict and International Humanitarian Law' (2012) 94 Intl Rev of Red Cross 515, 519.

[32] Toomas Hendrik Ilves, 'Address' (62nd Session of the United Nations General Assembly, New York, 25 September 2007) <www.un.org/webcast/ga/62/2007/pdfs/estonia-eng.pdf> accessed 10 January 2019. Other states have similarly referred to cyber-attacks as an emerging weapon for conflict, see Marco Roscini, 'Cyber Operations as a Use of force' in Nicholas Tsagourias and Russell Buchan (eds) *Research Handbook on International Law and Cyberspace* (Edward Edgar Publishing 2015) 239.

[33] Pavel Polityuk, 'Ukraine sees Russian hand in cyber attacks on power grid' (*Reuters*, 12 February 2016) <www.reuters.com/article/us-ukraine-cybersecurity/ukraine-sees-russian-hand-in-cyber-attacks-on-power-grid-idUSKCN0VL18E> accessed 21 February 2019.

weapon of attack in its recent invasion of Ukraine.[34] These incidences are only to mention a few. Many states and scholarly discussions now recognised that cyber-attacks can indeed be used as a weapon of war against the territorial integrity and political independence of states.[35] It has thus been advanced that the inclusion of the 'catch-all phrase', 'in any other manner inconsistent with the purposes of the United Nations',[36] persuasively connotes an inclusion of all the varying grave forms of conducts against states that are or should be prohibited as armed force.[37] Article 2(4) should be interpreted in such a way that essentially captures

> each and every form of armed force by individual States. Only such an extensive reading is apt to accommodate the universally acknowledged fundamental character of the prohibition and to fulfil the relevant purpose of the UN Charter, which is to ensure that armed force should generally not be used by individual states against each other.[38]

Cyber-attacks can manifest different results and their impact could be immediate or delayed.[39] A cyber-attack may not manifest tangible consequences but this does not in any way render it less of a cyber-attack.[40] A primary effect can be the alteration or deletion of

---

[34] Joe Tidy, 'Ukraine crisis: "Wiper" discovered in latest cyber-attacks' (*BBC News*, 24 February 2022) <www.bbc.co.uk/news/technology-60500618> accessed 08 March 2022.

[35] Stephenie G Handler, 'The new cyber face of battle: developing a legal approach to accommodate emerging trends in warfare' (2012) 48 Stan J of Intl L 209.

[36] UN Charter (n 23) art 2(4).

[37] It has been stated that the intention of the drafters in inserting this phrase was to leave no room for 'loopholes': Christian Henderson, *The Use of Force and International Law* (CUP 2018) 21;
Dorr (n 25) para 14;
Thomas Franck, *Recourse to Force: State Action Against Threats and Armed Attacks* (CUP 2002) 12-13;
Christian Tams, 'The Use of Force against Terrorists' (2009) 20 EJIL 364.

[38] Dorr (n 25) para 13.

[39] Marco Roscini, *Cyber Operations and the Use of Force in International Law* (OUP 2014) 52-53.

[40] It could be cybercrime, cyber espionage, a calculated cyber-attack or be all three, with internal or external effects. Irène Couzigou, 'Securing cyber space: the obligation of States to prevent harmful international cyber operations' (2018) Intl Rev of Law, Computers & Tech 37, 38.

relevant data or software on targeted computer systems. A secondary one can be the temporary or permanent incapacitation of infrastructure which the computer systems or networks service. A more substantial effect may be the destruction of property or even the loss of life. A cyber-attack may result in any or all of the above. It is specifically because of their capacity to result in substantial damage that some theories have been formulated to determine how and when a cyber-attack can actually be deemed a use of force under international law.[41]

Instrument-based or Means Approach

This theory focuses on armed attacks from the perspective of the weapons used in perpetrating them. This argument hinges on the fact that the drafters of the UN Charter premised article 2(4) on the need to prohibit arbitrary use of force by kinetic means, which was the prevalent method of warfare at the time. Such weapons were capable of producing any explosive effect 'with shock waves and heat'.[42] However, this notion was criticised on grounds that it would be excluding weapons that are biological, chemical or nuclear in nature.[43] Although these weapons are not always capable of producing shock waves and heat, they have the capacity to cause destruction to lives and properties, thus conforming with the notion of the use of force.[44] This argument has come to be accepted under international law. So, the instrument-based approach emphasises the tool used in actualising the force.[45] Essentially, the weapon used in an attack must be of a kinetic nature in order for the attack to qualify as an armed attack.[46] This approach has been long standing, being utilised to distinguish kinetic attacks from other economic or political

[41] Handler (n 35) 210; Buchan (n 16) 213-214.
[42] Brownlie (n 24) 362.
[43] ibid.
[44] ibid; Russell Buchan and Nicholas Tsagourias, *Regulating the Use of Force in International Law: Stability and change* (Edward Elgar Publishing 2021) 21.
[45] Irène Couzigou, 'The Challenges Posed by Cyberattacks to the Law on Self-defence' in August Reinisch, Mary E Footer and Christina Binder (eds), *International Law and…,* (Hart Publishing 2016) 250.
[46] Ido Kilovaty, 'Cyber Warfare and the Jus ad Bellum Challenges: Evaluation in Light of the Tallinn Manual on the International Law Applicable to Cyber Warfare' (2014) 5 National Security Law Brief 91, 109.

upheavals.[47] It has thus been advanced that cyber-attacks may not rise to the level of armed force since it lacks the necessary tangibility of conventional weapons.

The crux of the issue is whether a cyber-attack, usually not characterised by physical weapons, can still be classified as a use of armed force. This takes into consideration the literal meaning of the word 'armed' which means 'equipped with a weapon'.[48] Since cyber-attacks are executed through cyberspace, the physical element of an instrument of force in military terms may indeed be lacking.[49] The conclusion is that a cyber-attack generally cannot amount to a use of force or an armed attack.

However, this can be contested on the basis that cyber-attacks are generally executed through networked computer routes, involving the moving of 'information between physical elements of cyber infrastructure'.[50] This would include the computer system itself and all its associated hardware, all constituting physical elements. A clarification of the words 'means and methods' has been issued to define weapons with reference to the manner in which they are used,[51] rather than in isolation. The 9/11 attacks, among other things, led to a sudden realisation among states that 'almost everything could be turned into a weapon, anytime'.[52] As aptly noted by the ICJ in the *Nuclear Weapons* case,[53] the provisions of the UN Charter on the use of force do not specify the kind of weapons that can be used to perpetrate

---

[47] Handler (n 35) 226.
[48] Roscini, *Cyber Operations and the Use of Force* (n 39) 49.
[49] Handler (n 35) 227.
[50] Nils Melzer, *Cyber Warfare and International Law* (UNIDIR Resources 2011) 5.
[51] Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (adopted 8 June 1977, entered into force 7 December 1978) 1125 UNTS 3, art 35, Commentary of 1987, [1402]. It is useful to also note the context in which this averment was made, which is humanitarian law, serving the purpose of mitigating human suffering and destruction, hence not directly bearing with the resort to force as such.
[52] Olaf Theiler, 'New threats: the cyber-dimension' (*Nato Review*, 4 September 2011) <www.nato.int/docu/review/2011/11-september/cyber-threads/en/index.htm> accessed 4 September 2018.
[53] *Legality of the threat or use of nuclear weapons* (Advisory Opinion) (1996) ICJ Rep 226 (*Nuclear Weapons* Case) para 39.

an armed attack. Therefore, these provisions apply to all uses of armed force irrespective of the weapons used. Conventional or otherwise, the weapon used in an attack is not itself relevant to assessing whether an operation qualifies as a use of force or an armed attack.[54] An armed attack may be perpetrated by 'conventional or unconventional, primitive or sophisticated, ordnance'.[55] 'The use of any device or number of devices which results in a considerable loss of life and/or extensive destruction of property must therefore be deemed to fulfil the conditions of an armed attack'.[56] Hence, the argument that a cyber-attack does not engage physical elements or weapons, lacks potency.

The instrument-based approach also premises that typical 'conflict' situations necessitate the use of armed force. Cyber operations, however, are not usually engaged for the purpose of 'conflict'.[57] In fact, cyber-attacks are the exceptional circumstances running contrary to how computer operations are normally built to function. Hence, a computer network or system is not typically a weapon. Even where computers are used to perpetrate cyber-crimes or steal sensitive government data, they are usually dealt with under various criminal laws adopted by states and generally fall outwith the ambit of the law on the use of force.[58] Still, under those exceptional circumstances where computer networks or systems are used to perpetrate cyber-attacks, there is the tendency for them to rise to the level of use of force, when the 'directness' and 'predictability of consequences' similar to a kinetic attack is considered.[59] One major view has postulated that where a cyber-attack completely blows out a power grid, it may be deemed a use of force because, prior to these

[54] M N Schmitt and L Vihul (eds), *Tallinn Manual 2.0 on the International Law applicable to Cyber Operations* (CUP 2017) (*Tallinn Manual 2.0*) 328.
[55] Dinstein (n 24) 221.
[56] Karl Zemanek, 'Armed Attack' in Rüdiger Wolfrum (ed) *Max Planck Encyclopedia of Public International Law (MPEPIL)* (OUP 2013) para 21.
[57] May Larry, 'The Nature of War and the Idea of "Cyberwar"' in Jens D Ohlin, Kevin Govern and Claire Finkelstein (eds) *Cyber War: Law and Ethics for Virtual Conflicts* (Oxford Scholarship Online 2015) ch 2.
[58] Daniel B Silver, 'Computer Network Attack as a Use of Force under Article 2(4) of the United Nations Charter' in Michael N Schmitt and Brian T O'Donnell (eds) *Computer Network Attack and International Law* (International Law Studies, Vol 76 2002) 99.
[59] Johann-Christoph Woltag, *Cyber Warfare: Military Cross-Border Computer Network Operations Under International Law* (Intersentia Publishing 2014) 143.

current developments in cyberspace, such an attack could have been perpetrated using some destructive form of kinetic force that would achieve the same result.[60]

Article 41 of the UN Charter,[61] authorises the Security Council to utilise 'measures not involving the use of armed force' in giving 'effect to its decisions'. The article includes the 'complete or partial interruption of… telegraphic, radio, and other means of communication' in its list of measures that will not constitute a use of armed force.[62] Cyber-attacks often attempt the complete or partial interruption of some means of communication. Some experts argue that this provision reflects the instrument-based approach and implies that cyber-attacks may not constitute a use of armed force.[63]

Under the instrument-based approach, there are cogent arguments for and against the contention that cyber-attacks fit into the traditional notion of the use of armed force. Some consider this theory 'outdated' and not applicable to the emerging challenges that cyber-attacks present.[64] Consequently, this theory does not present the best approach to determining whether or not a cyber-attack constitutes a use of force.

Target-based or Strict Liability Approach

This theory places emphasis on state infrastructures as the target of a cyber-attack. Under the target-based theory,

> cyber-attacks against critical infrastructure are automatically treated as armed attacks, due to the severe consequences that can result from disabling those systems.[65]

Critical infrastructure may include cyber infrastructure comprising data and information storage and management systems,

---

[60] David E Graham, 'Cyber Threats and the Law of War' (2010) 4 J of Nat Sec L & Policy 87, 91
[61] UN Charter (n 23) art 41.
[62] ibid.
[63] Handler (n 35) 227.
[64] Kilovaty (n 46) 110.
[65] Jeffrey Carr, *Inside Cyber Warfare* (2nd edn, O'Reilley Media Inc 2011) ch 4.

computing systems, networks and network grids located on the territories of states that improve state productivity and ensure interconnectedness.[66] It may also include services such as 'food, water, transportation, banking and finance, health, energy, governmental and public services'.[67] It has been opined that any intrusion into a state's critical infrastructure should be deemed a 'demonstration of hostile intent' due to this infrastructure's fundamental role in the functioning of the modern state.[68] Proponents of this view further contend that where such an attack destabilises such infrastructures to the extent of disrupting or paralysing the systems, the attack could be considered an armed attack even if no immediate injury or material damage is recorded.[69] Such intrusion should trigger an immediate right to self-defence under article 51 of the UN Charter.[70]

Proponents further argue that even a cyber-attack perpetrated against a state's financial institutions resulting in severe economic loss for the state constitutes an attack against critical state infrastructure that should be deemed an armed attack.[71] A cyber-attack that destabilises the command and control systems belonging to a state's military may qualify as an armed attack against the state because the cyber-attack would be disabling a state's crucial defence and security infrastructure.[72] Accordingly, the Stuxnet attack on Iranian nuclear facilities in 2010 was a target-based cyber-attack. Stuxnet, a malware in form of a computer worm, was introduced through one vulnerable system that eventually spread across several other systems at a uranium enrichment plant. It replicated itself across the network

---

[66] Craig A Stewart, Stephen Simms, Beth Plale and others, 'What is Cyberinfrastructure?' (38th Annual ACM SIGUCCS Fall Conference, Virginia, October 2010) <http://dsc.soic.indiana.edu/publications/fp109a-stewart.pdf> accessed 24 May 2018.

[67] Nicholas Tsagourias, 'Cyber Attacks, Self-Defence and the Problem of Attribution' (2012) 17 J of Conflict & Sec Law 229, 231.

[68] Walter Gary Sharp, *Cyberspace and the Use of Force* (Aegis Research Corporation 1999) 130;
Christopher Bronk and Eneken Tikk-Ringas, 'The Cyber Attack on Saudi Aramco' (2013) 55(2) Survival: Global politics and Strategy 81, 90.

[69] Eric T Jensen, 'Computer Attacks on Critical National Infrastructure: A Use of Force Invoking the Right of Self-Defense' (2002) 38 Stan J Intl L 207, 222-223.

[70] Sharp (n 68) 130

[71] Tsagourias (n 67) 231.

[72] ibid.

systems in order to access information which was then used to destroy several centrifuges,[73] causing severe damage.[74] Some reports claim that the damage caused by the Stuxnet attack may have set back Iran's nuclear programme back by two years.[75] The US and Israel were allegedly behind these attacks.[76]

Target-based theorists have argued that the Stuxnet attacks constitute a use of force or even an armed attack perpetrated against Iran since the nuclear facility is a primary fuel enrichment plant.[77] There have been claims that the nuclear facility was particularly targeted as a medium of crippling Iran's development of nuclear weapons[78] even though the Iranian government has maintained that the nuclear plant was developed solely for peaceful purposes.[79] In the event, neither Iran nor other states have categorised the Stuxnet attack as a use of force, armed attack or even an act of aggression.[80] Iran was

[73] Centrifuges have been described as machine-controlling software that works by spinning materials at high speeds to separate out their components. The centrifuges at the Natanz facility worked to separate different kinds of uranium in order to specifically set aside the one needed for nuclear power and weapons. Gordon Corera, 'Iran Nuclear Attack: Mystery Surrounds Nuclear Sabotage at Natanz' (*BBC News*, 12 April 2021) <www.bbc.co.uk/news/world-middle-east-56722181> accessed 26 April 2021.

[74] Marie Baezner and Patrice Robin, *Cyber Defense Project: Hotspot Analysis: Stuxnet* (Centre for Security Studies 2017) Centre for Security Studies.

[75] David P Fidler, 'Cyberattacks and International Human Rights Law' in S Casey-Maslen (ed) *Weapons Under International Human Rights Law* (CUP 2014) 316; Stuxnet: targeting Iran's nuclear programme (2011) 17(2) Strategic Comments 1, 3.

[76] Lin (n 31) 519;
David E Sanger, 'Obama Order Sped Up Wave of Cyberattacks against Iran' *New York Times* (New York, 1 June 2012) <www.nytimes.com/2012/06/01/world/middleeast/obama-ordered-wave-of-cyberattacks-against-iran.html> accessed 19 February 2019.

[77] 'Natanz Enrichment Complex' (*NTI*) <www.nti.org/learn/facilities/170/> accessed 5 February 2019.

[78] It has been advanced that the Stuxnet worm locates specific targets and then wrecks damage on only those targeted network systems in a way that kinetic weapons may not be able to: Jeremy Richmond, 'Evolving Battlefield: Does Stuxnet Demonstrate a Need for Modifications to the Law of Armed Conflict?' (2012) 35 Fordham Intl L J 842, 852.

[79] There seems to be some concern within the international community that any development of nuclear materials will be geared towards the production of weapons of mass destruction: Buchan (n 16) 219.

[80] Fidler (n 75) 316;

not forthcoming with information on the extent of the damage caused by the Stuxnet attack.[81] Initial assertions surrounding the attack were deciphered from reports of the Institute for Science and International Security.[82]

Although opinion is not settled, there appears to be some support for the target-based approach within the international community.[83] The target-based approach may be an appropriate theory for critically analysing the classification of cyber-attack targets. However, the approach threatens the international law principles of the prohibition of the use of force and the limitation of state responses to the use of force. There is also no universally accepted definition of critical infrastructures, hence states may react to cyber-attacks against their critical infrastructures based on their own interpretations.[84] If any cyber-attack on a state's infrastructure is enough to trigger a reaction in self-defence, even if the attack does not register disastrous effects on lives and properties, the risk of abuse by states would be extremely high.[85]

James P Farwell and R Rohozinski, 'Stuxnet and the Future of Cyber War' (2011) 53 Survival 23, 29. But the Group of Experts on the Tallinn manual arrived at a different conclusion that it may constitute a use of force: M N Schmitt (ed) *Tallinn Manual on the International Law Applicable* to *Cyber Warfare* (CUP 2013) (Tallinn Manual) 58.

[81] The then head of Iran's Atomic Energy Organization, Ali Salehi, had commented that they were able to discover and halt the virus just in time before it could penetrate their equipment. Buchan (n 16) 219;
Farwell and Rohozinski (n 80) 29.

[82] David Albright, Paul Brannan, and Christina Walrond, 'Did Stuxnet Take Out 1,000 Centrifuges at the Natanz Enrichment Plant?' (I*nstitute for Science and International Security*, 22 December 2010) < http://isis-online.org/uploads/isis-reports/documents/stuxnet_FEP_22Dec2010.pdf> accessed 4 February 2018.

[83] Couzigou, *The Challenges Posed by Cyber Attacks* (n 45) 252.

[84] 'When do cyber operations amount to use of force and armed attack, and what response will they justify?' (*University of Oslo*, 2016) 22 <www.duo.uio.no/bitstream/handle/10852/50840/723.pdf?sequence=1> accessed 12 July 2018.

[85] Janne Valo, 'Cyber Attacks and the Use of Force in International Law' (Master's thesis, University of Helsinki 2014) 39.

Consequence-based or Effects Approach

This theory is the most widely accepted within the international community on when a cyber-attack can be considered a use of force or even an armed attack. The consequence-based approach places emphasis on the consequences of an attack as well as the severity of those consequences rather than the process by which it was carried out or which infrastructure was targeted. The International Court of Justice (ICJ), in the famous *Nicaragua* case, established that the use of force is of two kinds: the less grave forms of the use of force and the 'most grave forms of the use of force'.[86]

The fundamental difference is measured by the 'gravity'; that is, the 'scale and effects' of the attack. The 'most grave' forms of the use of force would essentially be 'relatively large scale', with 'sufficient gravity', capable of having 'substantial effect'.[87] Ascertaining the 'scale' of the attack comprises assessing where the attack was carried out, the duration and the amount of force used. Determining whether there has been a 'substantial effect' requires an assessment of the damage or destruction caused.[88] To establish 'most grave' uses of force, the damage or destruction should include cause severe damage to property or infrastructure, injury to life or limb, or death. 'Most grave' forms of the use of armed force are categorised as armed attacks under article 51 of the UN Charter.

Conversely, any use of force falling below the threshold of 'most grave' is deemed a less grave use of armed force. Examples of this may include intervention in the affairs of a state or the threat of force.[89] While states like Germany align with the ICJ's threshold of

---

[86] *Nicaragua* Case (n 25) para 191. The Court has reiterated this view in several other cases: *Nuclear Weapons* Case (n 53);
Armed Activities Case (n 25);
Oil Platforms (Islamic Republic of Iran v United States of America) (Judgement) [2003] ICJ Rep 161.

[87] Georg Nolte and Albrecht Randelzhofer, 'Article 51' in Bruno Simma, Daniel-Erasmus Khan, Georg Nolte and others (eds) *The Charter of the United Nations: A Commentary*, vol 2 (3rd edn, OUP 2012).

[88] Tom Ruys, '*Armed Attack' and Article 51 of the UN Charter: Evolutions in Customary Law and Practice* (CUP 2010) 140.

[89] *Nicaragua* Case (n 25) para 195.

gravity and severity to distinguish a lesser use of force from an armed attack,[90] states such as the US generally regard any use of armed force as an armed attack.[91]

With cyber-attacks, the key factor is whether the 'overall effects of a computer network attack are equivalent to those of an attack using a traditional weapon'.[92] Some cyber-attacks may not result in any physical damage; for example, those perpetrated to manipulate information on computer systems or cause disruption on government websites.[93] In the same vein, some cyber-attacks may be psychological, non-destructive and calculated to undermine the economy or government of a state. These kinds of disruptions may not be deemed a use of force or an armed attack seeing as their consequences are not generally serious or catastrophic.[94] However, cyber-attacks that cause serious disastrous effects or damage such as destruction of property, serious injury or death could be regarded as a most grave form of the use of armed force.[95] Examples could include: manipulation of air or railway controls resulting in a serious accidents; disruption of the control systems to nuclear power plants resulting in the release of harmful material; disabling the controls for hospital electricity leading to the shutdown of all emergency equipment and machines; or the disabling of controls for water systems and dams.[96]

---

[90] Norbert Riedel, 'Cyber Security as a Dimension of Security Policy' (speech at Chatham House, London, 18 May 2015) <www.auswaertiges-amt.de/en/newsroom/news/150518-ca-b-chatham-house/271832> accessed 20 December 2021;
The Tallinn Manual also adopts this view: *Tallinn Manual 2.0* (n 54) r 71, paras 6-8.

[91] Harold Koh, 'International Law in Cyberspace' (USCYBERCOM Inter-Agency Legal Conference, Maryland, September 2012) 4 <https://digitalcommons.law.yale.edu/cgi/viewcontent.cgi?article=5858&context=fss_papers> accessed 12 December 2019;
The Chatham House Principles also aligns with the view adopted by the US: Chatham House, *Principles of International Law on the Use of Force by States in Self-Defence* (The Royal Institute of International Affairs 2005) principle 2.

[92] Couzigou, *The Challenges Posed by Cyber Attacks* (n 45) 250.

[93] Buchan (n 16) 212.

[94] *Tallinn Manual* (n 80) 46-48.

[95] Dorr (n 25) para 12; Couzigou, *The Challenges Posed by Cyber Attacks* (n 45) 251.

[96] Couzigou, *The Challenges Posed by Cyber Attacks* (n 45) 250; Koh (n 91) 4.

The International Group of Experts on the Tallinn Manual have also affirmed that cyber-attacks 'that injure or kill persons or physically damage or destroy objects' are considered a use of force,[97] and may likely amount to an armed attack. The Experts adopted some guiding factors that could aid in ascertaining whether a cyber-attack constitutes a use of force. They include: the severity of the cyber-attack; the immediacy of the attack that may necessitate a response; the directness of the act, invasiveness; whether the effects of the cyber-attack are measurable; whether the cyber-attack itself is legitimately permissible under international law; and whether the cyber-attack will incur any state responsibility.[98] The Group of Experts were careful to conclude that these factors do not represent binding legal criteria for ascertaining whether a cyber-attack is an armed attack. They are mere factors in guiding states' assessments and do not constitute any rule under international law.

There have been cyber operations perpetrated in recent times that have triggered the likely application of the consequence-based theory in deciphering whether the attacks constitute a use of force or armed attacks under international law. Attacks on Estonia in 2007, for example, targeted various websites and their servers, including those of the government. Some banking systems were also disrupted. The attack was systematically organised and aimed at a Distributed Denial of Service (DDoS). The origin of the attack was difficult to ascertain, though Russia was suspected.[99]

Although the attack has been cited as an example of harmful cyber operations, it was never declared to be a use of armed force or an armed attack because it did not cause injuries or death. Neither did it result in any severe physical damage asides the internet traffic jam

---

[97] *Tallinn Manual 2.0* (n 54) 333.

[98] *Tallinn Manual* (n 80) 48. These factors were initially propounded by Schmitt in his earlier article: Michael N Schmitt, 'Computer Network Attacks and the Use of Force in International Law: Thoughts on a Normative framework (1999) 37 Columbia J of Trans L 885, 914-915;
Woltag (n 59) 145.

[99] Ian Traynor, 'Russia accused of unleashing cyberwar to disable Estonia' *The Guardian* (London, 17 May 2007) <www.theguardian.com/world/2007/may/17/topstories3.russia> accessed 2 February 2018.

and disruption of some banking operations.[100] Some remarks suggest that the attack could be deemed a lesser form of the use of force.

The WannaCry malware attacks disrupted the communication systems of multiple banks, hospitals and companies in May 2017.[101] At the time, it was described as the largest ransomware attack to have ever occurred. The British NHS was one of the major organisations affected as the cyber-attack, which rendered several health facilities inaccessible. Despite the damage caused, the WannaCry attack was not identified as a use of force or even an armed attack against the affected states since it did not result in major disasters or deaths or any physical destruction of operating structures. Hence, it was termed a low-intensity cyber-attack. If the cyber-attack had destroyed major hospital equipment to the extent of rendering them inaccessible to patients in dire need of them, resulting in loss of life, the WannaCry attack could have been deemed a use of force, even rising to the level of an armed attack.

Fortunately, no cyber-attack to date has been declared a use of armed force or armed attack.[102] Nevertheless, from the foregoing, it is clear that cyber-attacks are capable of rising to the level of the use of force and are thus prohibited under article 2(4) of the UN Charter.[103] The prohibition of the use of force by states entails an outright ban on the use of armed force, in whatever form, by or on behalf of a state, against another state's territorial integrity or political independence. A cyber-attack perpetrated by a state or non-state actor against another state, through compromised computer networks or systems, can be considered a use of force where its scale and effects are similar to those of a kinetic attack declared as a use of force.[104] Where the scale

[100]The same conclusion was reached for the cyber-attacks in Georgia in 2008. See Tsagourias (n 67) 232.
[101] 'Cyber-attack: US and UK blame North Korea for WannaCry' (*BBC News*, 19 December 2017) <www.bbc.co.uk/news/world-us-canada-42407488> accessed 22 January 2018.
[102] Ian Y Liu, 'State Responsibility and Cyber Attacks: Defining Due Diligence Obligations' (2017) 4 Indonesian J of Intl & Comp L 191, 195.
[103] Buchan (n 16) 212.
[104] *Tallinn Manual 2. 0* (n 19) r 69

and effects are so substantial, exceeding the threshold for the use of force, a cyber-attack can be deemed an armed attack.[105]

This restrictive but inclusive interpretation is consistent with the purposes of the United Nations, especially in the light of evolving contemporary issues under international law. It challenges the mainstream interpretation of article 2(4) of the UN Charter and the extent of its application. After all, the UN Charter represents a 'living, growing system of rules' which can adapt to the needs of the international community based on how customary practice evolves.[106] Some experts have remarked that it is only a matter of time before a cyber-attack erupts into full-scale war.[107] A representative of the US Cyber Command had commented that 'the next war will begin in cyberspace'.[108] Cyber-attacks may represent be the new face of force in international law. The possibilities of an actual cyber-attack rising to the level of an armed attack may not be so far-fetched.

Conclusion

The use of armed force is prohibited under international law. While this interpretation has for a long time been interpreted to connote that force perpetrated by kinetic means is prohibited, there is an ongoing shift as to what the prohibition entails. The proliferation of cyber-attacks in diverse forms have invoked questions relating to the extent to which current international law is capable of intervening to mitigate its harmful effects. There appears to be a general consensus that international law does apply to cyberspace operations. However, with regards to the law on the use of force, particularly whether cyber-attacks have the capacity of being equated with the conventional notion of armed attacks, the matter remains somewhat unsettled.

---

[105] Buchan and Tsagourias (n 44) 22, 123; *Tallinn Manual 2. 0* (n 19) r 71 (6).
[106] Thomas Franck as cited in Ruys, *The Meaning of 'Force'* (n 24) 163.
[107] Matthew C Waxman, 'Cyber-Attacks and the Use of Force: Back to the Future of Article 2(4)' (2011) 36 YJIL 421, 423-424;
Yosef Harash, 'The Next War will Happen in the Cybersphere, and Here's What it Will Look Like' (*Haaretz*, 6 July 2019) <www.haaretz.com/us-news/.premium-what-a-future-cyberwar-will-look-like-1.7441901> accessed 17 September 2019.
[108] Rex Hughes, 'A Treaty for Cyberspace' (2010) 86 International Affairs: Royal Institute of International Affairs 523.

Major arguments tilt towards the notion that cyber-attacks can also qualify as a use of armed force. The theories underpinning this school of thought are the instrument-based, target-based and the consequence-based approaches. The instrument-based approach largely defines a use of force by reference to the weapon used to perpetrate the attack. The target-based approach advances that any attack that targets critical infrastructure is deemed a use of force. The consequence-based approach identifies a cyber-attack as a use of force based on the gravity of its effects. None of these approaches are immune from criticism in terms of their application and practicality. However, of the three theories, the consequence-based approach appears to have gathered more recognition since it steadily aligns with the ICJ's description of a use of armed force and armed attack. It also aligns with the fundamental purpose of the UN Charter, which is to keep the scourge of war at bay. Hence, a cyber-attack is capable of being classified as a use of armed force under the UN Charter, depending on its gravity. Where a cyber-attack is perpetrated with such gravity that it results in severe destruction of property or severe injury or death, it can be classified as a severe use of armed force amounting to an armed attack under international law.

The Better Business Act: Putting the 'Enlightened' in Enlightened Shareholder Value?

EUAN B REID*

Abstract

Section 172 of the Companies Act 2006 embeds the concept of enlightened shareholder value in UK corporate governance. Yet this provision does little to deviate from the traditionally shareholder-centric nature of company law. In recognition of this fact, various businesses have come together to lobby for reform of section 172. However, the effectiveness of any Better Business Act will likely be limited in the absence of fundamental changes in the structure of corporate governance.

Keywords: corporate governance, company law, directors' duties

Introduction

'All thoughtful people believe that corporate enterprise should be organized and operated to serve the interests of society as a whole'.[1]

The way reformed UK company law strives for this aim is through Enlightened Shareholder Value ('ESV'). Yet ESV does very little to reassure those concerned by the fact that non-shareholder stakeholders may bear residual risk. However, a recent business-led initiative, the Better Business Act ('BBA'), seeks to drastically change this situation by placing a greater emphasis on non-shareholder interests through an amendment to section 172 of the Companies Act 2006. This article will consider the concept of ESV, its embodiment in UK company law and corporate governance, the BBA initiative and its likely impact if implemented.

*DPLP student, University of Aberdeen.

[1] Henry Hansmann and Reinier Kraakman, 'The End of History for Corporate Law' (2001) 89 Geo LJ 439, 441.

Background

What is the purpose of the company? One theory is shareholder primacy. According to this view, the goal of corporate officers is 'to make as much money for their stockholders as possible.'[2] The theory is based on the premise that as shareholders are residual claimants of companies they ultimately bear the risk that the company fails.[3] Accordingly, the one social responsibility of companies is 'to use its resources and engage in activities designed to increase its profits so long as it stays within the rules of the game'.[4]

However, the shareholder-management relationship 'is not the only relationship … important to wealth creation.'[5] Shareholders are rarely solitary residual risk-bearers. Employees with firm-specific knowledge and skills are a key example.[6] This gives rise to the stakeholder theory of the company: that companies should 'be managed in the interests of a broader range of stakeholders, including employees, consumers, and even the public at large.'[7]

When initiating the review of UK company law in the late 1990s, the New Labour government was conscious of this debate. The project's terms of reference included reforming the law to permit 'the maximum amount of freedom and flexibility to those organising and directing the enterprise' whilst protecting 'the interests of those involved with the enterprise, including shareholders, creditors and employees'.[8]

---

[2] Milton Friedman, *Capitalism and Freedom* (University of Chicago Press 1962) 133.
[3] Andrew Keay, *The Enlightened Shareholder Value Principle and Corporate Governance* (Routledge 2013) 18.
[4] Friedman (n 2) 133.
[5] Margaret M Blair, 'For Whom Should Corporations Be Run?: An Economic Rationale for Stakeholder Management' (1998) 31(2) Long Range Planning 195, 199.
[6] ibid 196.
[7] Virginia Harper Ho, '"Enlightened Shareholder Value": Corporate Governance Beyond the Shareholder-Stakeholder Divide' (2010) 36(1) Journal of Corporation Law 59, 71.
[8] Department of Trade and Industry (DTI), *Modern Law for a Competitive Economy* (DTI 1998) para 5.2.

The Company Law Review Steering Group acknowledged the present (shareholder primacy) model could potentially give rise to an unhelpful focus on short-term profits, and that arguably 'shareholders are [no longer] the sole repositories of residual risk'.[9] Two approaches to reform were therefore identified: 'pluralism', analogous to stakeholder theory,[10] and ESV, which maintains increasing shareholder value as an overarching aim but acknowledges that this is best achieved by considering a wider range of interests.[11] Exclusively focusing on profits is often detrimental to cultivating long-term relationships, which may involve short-term costs but provide longer-term benefits.[12] The Steering Group opined that while directors would often take an inclusive, long-term approach, the law was often misunderstood.[13]

Eventually, pluralism was rejected.[14] The Steering Group concluded that pluralism would confer too wide a discretion on directors, making them effectively unaccountable.[15] The *objective* of company law should be pluralist, but this 'should not be done at the expense of turning company directors from business decision makers into moral, political or economic arbiters'.[16] The government duly endorsed ESV and committed to embed the concept in its reforms under the Companies Act 2006.[17]

The Companies Act 2006

Section 172(1) of the 2006 Act provides that a director 'must act in the way he considers, in good faith, would be most likely to promote the success of the company for the benefit of its members as a whole'. When doing so they must 'have regard' to a non-exhaustive list of

[9] CLRSG, *Modern Law for a Competitive Economy: The Strategic Framework* (CLRSG 1999) para 5.1.10.
[10] ibid para 5.1.13-5.1.14.
[11] ibid para 5.1.12.
[12] ibid.
[13] ibid paras 5.1.19-5.1.20.
[14] CLRSG, *Modern Company Law For a Competitive Economy: Developing the Framework* (CLRSG 2000) para 3.20-3.31.
[15] ibid para 3.24.
[16] ibid para 2.21.
[17] DTI, *Company Law Reform* (Cmnd 6456, 2005) 5.

factors including employee interests and environmental and community impact.[18] In addition, section 172(3) states the duty 'has effect subject to any enactment or rule of law requiring directors, in certain circumstances, to consider or act in the interests of creditors of the company.' This recognises legal requirements for directors to take account of creditor interests during insolvency.[19]

The Secretary of State for Trade and Industry outlined the government had created a scheme which was 'workable', allowing directors to regard wider interests whilst protecting them from litigation.[20] However, although workable for directors, ESV offers little to non-shareholder stakeholders.

Continuing Shareholder Focus

At first non-shareholder stakeholders seem safeguarded. Directors must 'have regard to' the factors listed in section 172(1)(a) to (f) when promoting 'the success of the company'. The explanatory notes to the 2006 Act state, '[i]t will not be sufficient to pay lip service to the factors' contained in s172(1).[21]

However, once the factors are 'regarded', it is seemingly open for directors to make decisions which are disastrous for non-shareholders as long as they promote the success of the company for the benefit of the members as a whole.[22] Characterising directorial decision making as a balancing of interests is therefore inaccurate: advancing shareholder interests is the overarching objective.[23] Moreover, a company's 'success will usually mean long-term increase in value.'[24] Therefore, under ESV, non-shareholder interests have

[18] 2006 Act, ss 172(1)(b) and (d).
[19] Elaine Lynch, 'Section 172: A Ground-Breaking Reform of Director's Duties, or the Emperor's New Clothes?' (2012) 33(7) Company Lawyer 196, 199-200.
[20] HC Deb Tuesday 6 June 2006, vol 447, cols 125-126 (Alistair Darling).
[21] Explanatory Notes to the Companies Act 2006, para 328.
[22] Andrew Keay, 'Section 172 of the Companies Act 2006: An Interpretation and Assessment' (2007) 28(4) Company Lawyer 106, 108.
[23] Paul Davies and Jonathan Rickford, 'An Introduction to the New UK Companies Act: Part 1' (2008) 5(1) ECFR 48, 66.
[24] DTI, *Companies Act 2006: Duties of Company Directors: Ministerial Statements* (DTI 2007) 7

value only while they coincide with the interests of the members;[25] are deserving of protection only while contributing to profits.

Poor Enforceability

The duty in section 172(1) is owed *to the company*.[26] Non-shareholders therefore have no direct right of action if they suspect directors have breached section 172.[27] This is perhaps surprising given section 309 of the Companies Act 1985, which required directors to regard employee interests when performing their functions, was deemed 'next to useless' for this reason.[28]

Yet shareholders can, subject to court approval,[29] bring derivative proceedings against directors in respect of a right of action vested in the company.[30] Non-shareholders must then look to members to protect their interests. However, while shareholders might feel sufficiently aggrieved to initiate proceedings where they feel directors have not acted fairly as between members[31] or promoted the success of the company, intervention on behalf of non-shareholders is unlikely given the costs involved.[32] While non-shareholder stakeholders could potentially purchase shares to avail themselves of this option,[33] procedural requirements may make such a tactic unlikely to succeed.[34]

Admittedly, while a member might have interests overlapping with a section 172(1) factor (e.g. as an employee) and bring derivative proceedings, this is ultimately self-interest in another capacity and is

<https://webarchive.nationalarchives.gov.uk/20090609024504/http://www.berr.gov.uk/files/file40139.pdf> accessed 28 December 2021.

[25] Keay, *Interpretation* (n 22) 108.

[26] 2006 Act s170(1).

[27] Keay, 'Interpretation' (n 22) 109.

[28] ibid.

[29] 2006 Act ss261(1), 266(1).

[30] 2006 Act ss260(1),(3), 265(1),(3).

[31] 2006 Act s172(1)(f).

[32] Keay, 'Interpretation' (n 22) 109.

[33] Keay, *Corporate Governance* (n 22) 140.

[34] Alistair Alcock, 'An Accidental Change to Directors Duties?' (2009) 30(12) Company Lawyer 362, 368.

unlikely to occur frequently.[35] Even if activist shareholders successfully influenced directorial decision making, this risks others litigating on the basis their share price has been adversely affected.[36] Indeed, the very ability of minority shareholders to bring derivative proceedings might make directors 'more likely to risk a technical breach of s 172(1) than fail to promote the interests of the company.'[37]

Yet the foregoing is likely irrelevant. If derivative proceedings were initiated a director can simply argue they – in good faith – regarded the section 172(1) factors and believed their actions would promote the success of the company.[38] Moreover, establishing that directors did not 'regard' the section 172(1) factors could prove extremely difficult.[39] The position of directors is therefore 'virtually unassailable',[40] with non-shareholder stakeholders left without means to protect their interests.

Inadequate Transparency

One of the Steering Group's guiding principles was 'that freedom of contract and exchange in the broadest sense, *supported by transparency requirements*, should be the approach wherever possible.'[41] Accordingly, listed companies would produce an Operating and Financial Review ('OFR') including 'all that is material in the directors' view for users to achieve a proper assessment of the performance and future plans and prospects of the business.'[42] This would include, where relevant, information regarding employee (and other) relationships, and environmental and community impact.[43]

---

[35] Andrew Keay, 'Tackling the Issue of the Corporate Objective: An Analysis of the United Kingdom's Enlightened Shareholder Value Approach' (2007) 29 Sydney LR 577, 609.
[36] *R (on the application of People & Planet) v HM Treasury* [2009] EWHC 3020 (Admin) [34] (Sales J).
[37] Deryn Fisher, 'The enlightened shareholder - leaving stakeholders in the dark: will section 172(1) of the Companies Act 2006 make directors consider the impact of their decisions on third parties?' (2009) 20(1) ICCLR 10, 16.
[38] Keay, *Corporate Objective* (n36) 609.
[39] Lynch (n 19) 201.
[40] Keay, 'Interpretation' (n 22) 108.
[41] CLRSG, *Strategic* (n 9) para 2.22 (emphasis added).
[42] CLRSG, *Developing* (n 14) para 2.19.
[43] ibid.

This 'would enable shareholders and the community as a whole to monitor performance by directors of the broadly expressed inclusive duty.'[44] That the OFR was integral to successfully implementing ESV is clear: the Steering Group described the inclusive duty and OFR as 'mutually reinforcing'.[45]

However, the OFR was abandoned before the 2006 Act was enacted.[46] Although the Business Review would later replicate many of the requirements of the OFR, Burns and Paterson emphasise that it is subject to a less stringent audit regime with only piecemeal guidance available.[47] This potentially leaves narrative reporting requirements vulnerable to abuse, further reducing any effectiveness of ESV in protecting non-shareholders.

Indirect Protection

Yet perhaps ESV can *indirectly* promote protection of non-shareholder stakeholders. Firstly, while seemingly possible to consider non-shareholder interests when acting in the interest of members at common law, there is now *duty* to do so.[48] This legitimises actions of directors who wish to consider other stakeholders (while promoting the success of the company for members).[49] Further, by explicitly highlighting non-shareholders, section 172(1) provides 'a legislative lesson in good management and corporate social responsibility.'[50] Alternatively fear (well-founded or not) of litigation may instil a culture of more inclusive directorial decision making.[51]

---

[44] ibid para 2.20.
[45] ibid para 2.22.
[46] See Tom Burns and John Paterson, 'Gold plating, gold standard or base metal? Making sense of narrative reporting after the repeal of the Operating and Financial Review Regulations' (2007) 18(8) ICCLR 247, 253ff.
[47] ibid 258-259.
[48] Davies and Rickford (n 23) 66.
[49] Keay, 'Corporate Objective' (n35) 599-601.
[50] Lynch (n 19) 202.
[51] Joan Loughrey, Andrew Keay and Luca Cerioni, 'Legal Practitioners, Enlightened Shareholder Value and the Shaping of Corporate Governance' (2008) 8(1) JCLS 79, 102.

However, socially conscious directors would *already* have considered non-shareholder stakeholders.[52] Meanwhile the aforementioned lack of enforceability of section 172(1) allows irresponsible directors to continue disregarding wider interests.[53] This assertion is strengthened by evidence indicating section 172(1) has had negligible impact on directorial decision making.[54] It remains to be seen whether recently introduced requirements for large companies to state in their strategic report how directors have regarded the section 172(1)(a) to (f) factors will result in perceptible change.[55]

A Surprising Result?

With New Labour's talk of the 'stakeholder economy, where everyone has a stake in society and owes responsibilities to it',[56] it seems odd that non-shareholder stakeholders are left so exposed under ESV. However, political context is important. Wen and Zhao outline how New Labour largely followed the liberal economic philosophy of preceding Conservative administrations, making continuance of shareholder-orientated corporate governance more likely.[57]

This is reflected in the remit given to the Steering Group. The overarching ethos of the Company Law Review was economic competitiveness and increasing prosperity, *not* protecting stakeholders.[58] The eventual outcome reached is perhaps not overly surprising. Protection for non-shareholders thus lies outwith

---

[52] Nicholas Grier, 'Enlightened Shareholder Value: Did Directors Deliver?' (2014) JR 95, 107.
[53] ibid.
[54] Samantha Fettiplace and Rebecca Addis, *Evaluation of the Companies Act 2006, Volume One* (Department for Business, Innovation and Skills 2010) 73, 161-2 <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/31655/10-1360-evaluation-companies-act-2006-volume-1.pdf> accessed 28 December 2021.
[55] Companies Act s414CZA.
[56] Labour Party, 'General Election Manifesto 1997' in Iain Dale (ed), *Labour Party General Election Manifestos 1900-1997* (Routledge 2000) 360.
[57] Shuangge Wen and Jingchen Zhao, 'Exploring the Rationale of Enlightened Shareholder Value in the Realm of UK Company Law - The Path Dependence Perspective' (2011) 14 Intl Trade & Bus L Rev 153, 169.
[58] DTI, *Modern Company Law for a Competitive Economy* (n 8) at 1.2.

corporate law (e.g. employment and environmental regulation),[59] exactly as argued by advocates of shareholder primacy.[60]

Those concerned that non-shareholder stakeholders may bear residual risk will be little reassured by ESV. The 2006 Act treats shareholder interests as overriding whilst giving no effective means for non-shareholder stakeholders to enforce section 172. This situation is worsened by the scrapping of the OFR. Non-stakeholders seeking protection must therefore look beyond corporate law.

The Better Business Act

ESV has come under renewed scrutiny of late. At the time of writing, more than 870 businesses have pledged their support to a Better Business Act ('BBA').[61] In short, the BBA entails an amendment of section 172 with a stated aim of legally obliging companies 'to operate in a manner that benefits their stakeholders, including workers, customers, communities and the environment, while seeking to deliver profits for shareholders.'[62] Accordingly, a draft text of the BBA has been produced by the campaign.[63] After amendment the new section 172 would require a director to 'act in the way the director considers, in good faith, would be most likely to advance the *purpose* of the company' whilst having regard to a non-exhaustive list of factors identical to the current formulation of section 172.[64]

This seems initially to be a minor modification of the status quo. However, the key change is that the purpose of the company will not *only* be to benefit the company's members as whole. Additionally,

---

[59] Lynch (n 19) 202-203.
[60] Hansmann and Kraakman (n 1) 442.
[61] A full list of supporters is available at <https://betterbusinessact.org/supporters/> accessed 28 December 2021.
[62] BBA, 'FAQ and Resources' < https://betterbusinessact.org/frequently-asked-questions/> accessed 28 December 2021.
[63] Available at <https://betterbusinessact.org/wp-content/uploads/2021/04/The-Better-Business-Act-2021.pdf> accessed 30 December 2021.
[64] BBA 2021 (emphasis added) <https://betterbusinessact.org/wp-content/uploads/2021/04/The-Better-Business-Act-2021.pdf> accessed 30 December 2021.

its purpose would be to operate in a manner which benefits wider society and the environment in a manner commensurate with the size of the company and the nature of its operations. The company should also reduce any harms it creates or costs it imposes on wider society or the environment, with the goal of eliminating any such harms or costs.[65] A company may specify a different purpose in its articles with the caveat that this must be '*more* beneficial to wider society and the environment'.[66] In addition, this change to section 172 would be supplemented by new reporting requirements.[67]

Business as Usual?

The BBA reforms entail a radical change to the UK corporate governance architecture. For instance, a working group from the Institute of Directors has described the BBA as 'empower[ing] directors to pursue regenerative principles while leaving them a wide margin of discretion in their decision-making on these issues.'[68]

In practice, however, as was the case before New Labour's reforms, ethical companies and directors already give adequate consideration to non-shareholder interests. It is arguable that the BBA proposals would reduce the risk of reprisal by shareholders against directors who they regard as giving undue consideration to non-shareholder interests, yet the converse is also true. Unscrupulous directors can continue to argue that they sought to advance members' interests in the mandated environmentally and socially conscious fashion in good faith and they will remain practically unimpeachable.

Furthermore, and perhaps most problematic for those seeking radical change, the BBA would not result in any change as to *who* can enforce directors' duties. Shareholders would retain their privileged position in this regard. However, other factors are at play. Environmental, social and governance (ESG) concerns and

[65] ibid.
[66] ibid (emphasis added).
[67] ibid.
[68] Regenerative Business Working Group, *Amending UK Company Law for a Regenerative Economy* (IOD 2021) 17 <https://betterbusinessact.org/wp-content/uploads/2021/05/IoD-CG-Centre-Amending-UK-Company-Law.pdf> accessed 28 December 2021.

responsible business practices appear to of growing importance to investors.[69] Given this trend, perhaps a growing class of altruistic shareholders would be able to effectively hold company directors to account through the BBA.

Yet even if shareholders are minded to engage with management on ESG issues, it should be remembered that shareholder activism is motivated by a range of factors.[70] A recent survey of asset managers indicated the most common reasons for engagement were strategy followed by financial performance.[71] Moreover, due to engagement costs asset managers prefer divesting shareholdings if unsatisfied with a company.[72] While wider factors may have increased in significance for some shareholders it is certainly not a given that these will trump economic considerations, reducing the positive impact of the BBA proposals, should they be implemented.

Additionally it is impossible to ignore the ever-increasing proportion of shares of UK public companies held by non-UK investors, which the Office for National Statistics reports reached a new peak of 54.9% of the value of the UK stock market in 2018.[73] Certainly, while recent publications from the Investment Association acknowledge the value of considering wider interests such as climate change,[74] this can hardly be said to be representative of attitudes of

---

[69] See BNP Paribas, *The ESG Global Survey 2021* (BNP Paribas 2021) 6, 11ff; Investment Association, *The Path to Net Zero: Investing in a carbon neutral future* (IA 2020) 9-10.

[70] Iris HY Chiu and Dionysia Katelouzou, 'Making a case for regulating institutional shareholders' corporate governance roles' (2018) 1 JBL 67, 73-74.

[71] Investment Association, *Stewardship in Practice: IA Stewardship Survey* (IA 2018) 11.

[72] Julia Mundy, Lisa Jack and Sandra Einig, 'Is the Stewardship Code Fit for Purpose?' (CIMA 2019) 2.

[73] Office for National Statistics, 'Ownership of UK Quoted Shares: 2018' (14 January 2020) <www.ons.gov.uk/economy/investmentspensionsandtrusts/bulletins/ownershipofukquotedshares/2018> accessed 28 December 2021. This trend seems to be continuing at pace – see Alison Owers, *Foreign investors take an increased stake in UK plc* (Orient Capital 2021) <https://www.orientcap.com/media/5nifl2bc/oc-foreign-investors.pdf> accessed 30 December 2021.

[74] Investment Association, *Good Stewardship Guide 2021* (IA 2021) 5.

the whole body of holders of UK equities. While domestic shareholders might feel sufficient political or social pressure to closely monitor the conduct of directors it is uncertain whether this attitude will permeate the minds of non-UK shareholders. Ultimately, even if the BBA proposals are implemented, it is by no means certain that shareholders will enforce it.

A working group from the Institute of Directors, when endorsing the BBA, described the 'necessary' amendment of section 172 precipitating a shift 'to an approach that embraces stakeholder governance'.[75] However, any movement towards stakeholdersim will likely be blunted. The BBA, while certainly providing a clear change in emphasis for directors, continues to place shareholders in the crucial monitoring role they currently occupy. While ESG might be an important agenda for some investors, this is hardly guaranteed.

Leaving aside the BBA, there are means by which true stakeholderism could be achieved. Blair, for instance, argues that all stakeholders who have 'firm-specific investments' at risk should be protected through contractual arrangements and governance mechanisms.[76] Where stakeholders bear significant risk such protections might entail compensation schemes and the allocation of meaningful numbers of shares to those groups.[77] The BBA proposes nothing so radical. Those who seek fundamental change to corporate governance practice are likely to left disappointed. ESV, while perhaps bruised, lives to fight another day.

Conclusion

It has been shown that ESV as embodied by the 2006 Act does little to deviate from shareholder-centric notions of corporate governance. It will remain to be seen whether the government will embrace the BBA and alter the formulation of section 172(1). However, even if implemented, the BBA reforms will likely not be sufficient to effect the shift in corporate governance culture which many regard as

[75] Regenerative Business Working Group (n 68) 15.
[76] Margaret M Blair, 'Rethinking Assumptions Behind Corporate Governance' (1995) 38(6) Challenge 12, 17.
[77] ibid.

necessary. The BBA will still rely on shareholders to monitor the managers of companies. If shareholders neglect that monitoring, or decide not to take adequate account of ESG factors, then other stakeholder groups would be left virtually unprotected. Despite lofty ambitions, it is unlikely that a BBA, in the form currently proposed, would guarantee more socially conscious directorial decision making.

Is Decommissioning Liability Creating a Barrier to Deal Activity on the United Kingdom Continental Shelf?

LAUREN WENDY THOMSON*

Abstract

The UK maintains the most expansive decommissioning bill of any country. The recent High Court case of Exxon v Apache *demonstrates the multitude of issues surrounding decommissioning liability following M&A activity that is beginning to arise before the courts. This paper examines decommissioning in the United Kingdom from a legal perspective, focusing particularly on the wide-ranging liability that exists under section 29 of the Petroleum Act 1998, which commentators have stated will be a rich seam for lawyers working in the sector as aspects continue to be challenged in court. This paper also analyses the issue of funding and the absence of law in this area, recommending that stricter legislation is implemented, with the possibility of an international agreement. It is concluded that liability under the Act is unarguably restrictive to M&A activity but that tax incentives and the role of private equity have contributed to mitigating this barrier.*

Keywords: decommissioning, Petroleum Act 1998, energy transition, oil, gas, residual liability, finance

Introduction

Theoretical discussion surrounding the legal and financial issues concerning the removal of offshore installations is no longer adequate. The United Kingdom's (UK) decommissioning bill will amount to the largest of any country. It is estimated that £16.6 billion is to be spent in the next decade[1] and as many as 470 installations are to be

* LLB, University of Aberdeen; LLM, BPP University. The author is currently a Trainee Solicitor at Clifford Chance.

[1] OGUK, *Decommissioning Insight* (OGUK 2021).

decommissioned in the North Sea over the next 30 to 40 years.[2] The recent high court case of *Apache UK Investments Limited v Esso E&P UK Ltd*[3] highlights the beginning of liability debates before the court, with commentators stating the case is 'setting a precedent for future debates around liability for decommissioning costs.'[4] The subject is no longer a matter for a later date, the time has arrived – is the industry, regulators, law and financial sector prepared?

Decommissioning is the process by which an operator of offshore oil and gas installations and pipelines plan for and shut down the operation of assets, which includes plugging wells, removing installations, and cleaning up the area for subsequent users of the sea. This event is triggered when the economics determine that the operation of a field (revenue less the operating costs) becomes negative, also known within the industry as 'cash flow negative'. When this stage occurs, production will be considered no longer economically viable, and decommissioning will be planned to commence within the following year or two. This is the typical methodology around abandonment within the upstream sector.

From a commercial perspective, the complexities and challenges that are associated with decommissioning offshore oil and gas assets have dampened deal activity in the past, however it will be explored and concluded that change has ensued, with multiple mature asset deals having been completed in recent years.[5] Particularly, there has been a significant increase in the number of UK Continental Shelf (UKCS) licence assignments and change of control

---

[2] Romana Adamcikova, 'UK North Sea Decommissioning: The £17 Billion Challenge - UK's Decommissioning Spend Predicted to Overtake Capex in 2025' (*Wood Mackenzie*, 24 February 2020) <www.woodmac.com/news/opinion/uk-north-sea-decommissioning-the-17-billion-challenge/> accessed 24 February 2020.

[3] *Apache UK Investment Ltd v Esso Exploration and Production UK Ltd* [2021] 4 WLR 85.

[4] Allister Thomas, 'Exxon loses to Apache in High Court Battle on £100m decommissioning bill' (*Energy Voice,* 2021) <www.energyvoice.com/oilandgas/north-sea/decom/324754/exxon-apache-decommissioning-court/> accessed 25 May 2021.

[5] David Baker, Tom Bartlett, Richard Jones and others, 'North Sea Decommissioning: Primed for a Boom?' (*White & Case,* 29 June 2017) <www.whitecase.com/publications/insight/north-sea-decommissioning-primed-boom> accessed 3 January 2020.

from large operators to smaller players, and an increase of asset transfers with more innovative agreements concerning the liability for decommissioning.[6]

There have been key concerns within the industry regarding the future life of companies and assets given the fluctuations in oil price. The M&A market has faced an all-time low over the last few years with deal count in January and February 2020 among the lowest this century.[7] However this has not always been the case and it is very likely that the industry will recover, particularly with asset sales to drive an increase of equity. The most apparent concern given the recent economic climate is that low oil prices for prolonged periods of time will impact assets economic viability, which will result in the need for fields to cease production and decommission earlier than anticipated. This makes decommissioning an even more prevalent subject to be explored and understood. Aside from the recent state of events, can decommissioning liability be held responsible for frustrations in deals?

This paper will explore four main themes – the regulatory regime surrounding decommissioning, tax regimes, private equity, and funding. Firstly, chapter one will investigate and take a microscopic look at the current legislation and regulations that are in place and assess its adequacy. This exploration will help the reader fully understand the severity of liability that is placed upon operators. It will continue to analyse the impact and the extent that liability can have on multiple parties associated with an installation. Chapter two will follow with a focus on the steps that the UK government have taken to mitigate the impact decommissioning liability is having to the commerciality of asset transfers, with a particular attention to the tax regimes in place where it will be concluded that fiscal regimes are paramount to the oil and gas industry's success and financial

[6] Department for Business, Energy & Industrial Strategy (BEIS), *Guidance Notes Decommissioning of Offshore Oil and Gas Installations and Pipelines* (BEIS 2018) 20. Hereinafter 'BEIS Guidance Notes'.

[7] Simon Flowers 'Coping With Negative Cash Flow: Can the M&A Market Help Oil Companies Bolster Finances' (*Wood Mackenzie*, 24 March 2020) <www.woodmac.com/news/the-edge/coping-with-negative-cash-flow/> accessed 24 March 2020.

attractiveness of deals. Chapter three will look at the developing role of private equity in deals and determine that they have been the answer to troubling times. The potential risk of liability to private equity funds will also be considered. Finally, funding will be the concluding topic and from the view of this paper, the most important area that industry and regulators need to address to ensure deals are concluded - although uncertainty over oil price is nothing new in this industry, given the current economic climate reduced access to financing is a new hurdle that needs to be overcome[8] and regulators ought to take account of this.

Generally, it will be concluded throughout this paper that decommissioning liability unarguably does place a barrier on the attractiveness of deals within the UKCS. Nevertheless, it will be recognised that the role the government has played in implementing new and revised fiscal regimes has commenced the removal of this barrier to making deals happen. It will however be argued that more work needs to be done to regulate the funding of decommissioning. Measures ought to be put in place as a matter of urgency, particularly with the prospect of early shut-ins resulting from fluctuating oil and gas prices over the last year.

Legislation Surrounding Decommissioning Liability

Regulation Overview

For one to fully appreciate the barrier that decommissioning liability can cause to deals within the UKCS it is imperative to explore and understand the magnitude of stringency that UK legislation, as well as national and international regulations place on operators to ensure that decommissioning is not only carried out in an appropriate manner, but that financial measures are put in place to prepare for such events. Recently there has been a sharpening of focus[9] on the way

[8] Nevian Boroujerdi, 'Oil Price Crash: Can the North Sea Survive at US$30/bbl?' (*Wood Mackenzie*, 22 March 2020) <https://my-legacy.woodmac.com/reports/upstream-oil-and-gas-oil-price-crash-can-the-north-sea-survive-at-us30bbl-395972?contentId=395972&source=21> accessed 22 March 2020.

[9] Marc Hammerson, *Upstream Oil and Gas: Cases, Materials and Commentary* (Globe Business Publishing Ltd 2011) 437.

decommissioning is regulated and carried out, with a significant emphasis on liability. Despite this fact, the legal issues that impend the subject in question have received limited attention by the energy law research community[10] and thus, it is critical to look at the current regimes in order to ascertain their effectiveness and identify any potential legal and business complications that may arise. The industry is entering a phase where energy infrastructure is aging coupled with a lack of experience on the magnitude and large-scale projects of redundant structures.[11] This will cause the need for a continuing evolution of regulation as the industry and regulators expand and develop their understanding of the matter. As will be seen, there are concerns regarding liability on both sides of a commercial transaction of an asset. The main concern for the seller is their continuing liability in conjunction with residual liability and the buyers concern is the extent of the liability transferred to them within the deal. These issues will be considered throughout this chapter.

The Road to the Current Regime

The obligation to decommission installations in the North Sea was first introduced by the Geneva Convention 1958 which required that any offshore installations must be 'entirely removed'.[12] This was reinforced by the United Nations Convention on the Law of the Sea 1982[13] requiring that installations are removed to the extent that is necessary 'to ensure safety of navigation, taking into account any generally accepted international standards'.[14] The 1996 Protocol to the London Convention 1972 shifted the obligation from permissive to restrictive.[15] The current most pressing international obligation is confirmed in the Oslo and Paris Commissions OSPAR Decision 98/3 on the Disposal of Disused Offshore Installations to which the UK must adhere to. This decision specifically prohibits the 'dumping and

[10] Raphael Heffron, 'Energy Law for Decommissioning in the Energy Sector in the 21st Century' (2018) 11 The Journal of World Energy Law & Business 189, 189.
[11] ibid.
[12] Article 5(5).
[13] Article 60(3).
[14] Article 60(3).
[15] Elizabeth A Kirk, 'The 1996 Protocol to the London Dumping Convention and the Brent Spar' (1997) 46(4) The International and Comparative Law Quarterly 957-964, 958.

the leaving wholly or partly in place of disused offshore installations within the maritime area',[16] with only limited exceptions to derogate.[17] This indicates that there is little room to avoid decommissioning liability under the constraints of international regulations and results in a multitude of considerations for operators, sellers, and buyers operating within the UKCS.

The international decision affirmed in OSPAR 98/3 in particular triggered the UK Government to introduce stricter regulations ensuring that the North Sea works towards a clean *'as found'* seabed. The Government has balanced its international obligations whilst also trying to incentivise investment in the UKCS by being cautious about creating disincentives to buyers with decommissioning liability obligations. What has been done? Has the Government incentives worked? What legislation in the UK ensures that liability will not fall to the taxpayer, and is it efficient?

At a domestic level, the UK Government enacted the Petroleum Act 1998 (1998 Act) which is the principal legislation governing decommissioning in the UK, specifically Part IV 'abandonment of offshore installations.' It is regulated by the Department for Business, Energy, and Industrial Strategy (BEIS) and the Offshore Petroleum Regulator for Environment and Decommissioning (OPRED). The 1998 Act is supported by Guidance Notes[18] produced by BEIS to provide further information to operators. The overall aim of the 1998 Act is to ensure that those who benefitted from the exploration or production of hydrocarbons bears the responsibility to decommission the asset,[19] whilst ensuring that the taxpayer is protected from funding the project. The 1998 Act has been amended, notably by the Energy Act 2008 which strengthened the financial powers of the Secretary of State, and the Energy Act 2016 which established and equipped the Oil and

[16] OSPAR Commission, *Ministerial Meeting of the OSPAR Commission, Sintra, 22-23 July 1998: Programmes and Measures* (OSPAR Commission 1998) 15, para 2. Hereinafter 'OSPAR Decision 98/3'.

[17] Steel installations weighing more than ten thousand tonnes in air; gravity based concrete installations; floating concrete installations; any concrete anchor-base which results, or is likely to result in interference with other legitimate users of the sea. See OSPAR Decision 98/3.

[18] BEIS Guidance Notes (n 4).

[19] ibid 10.

Gas Authority (OGA) with the power to maximise economic recovery of offshore petroleum licenses.[20] A focal point for the purpose of this paper is that the 2008 Act amended Part IV of the 1998 Act to account for changes in practices, particularly the increased participation of less prominent operating companies with smaller assets.[21]

Liability under the Petroleum Act 1998

The liability specified within the 1998 Act is wide-ranging, with the Secretary of State permitted to serve a Section 29 notice (which fixes an obligation on the holder to submit a decommissioning programme)[22] on virtually anyone who has been connected (or still is) with an offshore installation. The list of persons who can be served a Section 29 notice is extensive and is contained in Section 30 of the Act - 'persons who may be required to submit programmes'. It extends to licensees who have transferred an interest,[23] limited liability companies,[24] and a person already carrying out activities on an offshore installation (or intends to in the future).[25] In addition to this, an extended liability is laid out in Section 34 of the Act which allows the Secretary of State to require a further class of persons to take responsibility for the cost of decommissioning. This is often referred to as the 'last resort powers' to pursue owners, licensees and operators,[26] as well as entities which has been 'associated' with them.[27] Once a notice has been served the party becomes joint and severally liable[28] for all decommissioning costs relating to the installation referenced in the notice.

[20] ibid.
[21] ibid 12.
[22] ibid 14.
[23] Petroleum Act 1998, s 30(1).
[24] Subsections 2(b) and (3) amended paragraphs (1)(e) and (2)(c) of s 30 from company to 'body corporate' to ensure that a limited liability partnership can be served with a section 29 notice as an associated party.
[25] Petroleum Act 1998, s 30(5)(b).
[26] BEIS Guidance Notes (n 4) 14-15.
[27] Petroleum Act 1998, s 30(2).
[28] BEIS Guidance Notes (n 4) 14.

Essentially, as a result of the 1998 Act, historic parties never escape liability until the asset has been fully decommissioned.[29] This liability is exceptionally concerning for sellers of assets and undoubtedly a contributing factor to the frustration of North Sea transactions.[30] It is a particular fear for majors selling to smaller operators with less financial capability, especially in today's market with smaller companies more than ever entering to '*sweat the fields near depletion*'[31] who are at a higher risk of defaulting, resulting in the liability falling due to the major. The buyer will undertake due diligence on the application of 'associated entity' to establish the entities within its corporate structure that could have liability.[32] The methodology behind this is to protect the UK taxpayer from funding the decommissioning after a company has made billions in years preceding the event. Thus, although this last resort power has never been used, the potential liability alongside the increasing number of less financially strong buyers has led to sellers requiring protection to guard against potential future defaults (which will be considered in detail in chapter three) and is considered a key consideration[33] for sellers.

Decommissioning costs affirmed in a decommissioning programme are not always certain and there runs a high risk of over expenditure. For example, the North-West Hutton decommissioning

---

[29] James Phillips and Rosie Lord, 'Legal Briefing: Projects, Energy and Natural Resources' (*The In-House Lawyer*, 2012) <www.inhouselawyer.co.uk> accessed 20 December 2019.

[30] PWC, 'M&A in the North Sea: Overview and Recent Deals' (*PWC*, 2016) slide 9 <https://higherlogicdownload.s3.amazonaws.com/SPE/a77592d6-ec9a-43b1-b57b-c7275fb91cb0/UploadedImages/SPE%20Past%20Event%20Presentation%20Downloads/2016%20Presentations/PwCNorthSeaM&A310616.pdf> accessed 15 May 2022.

[31] James Phillips and Rosie Lord (n 29).

[32] Hywel Davies and Oliver Moir, 'Offshore Decommissioning Liability: "Hotel California" and Cross-Contamination Risks' (*Slaughter and May,* June 2019) <www.slaughterandmay.com/media/2537531/offshore-decommissioning-liability-hotel-california-and-cross-contamination-risks.pdf> accessed 9 March 2020.

[33] Davies and Moir (n 32).

programme estimated a cost of £160m[34] but resulted in costing £246m.[35] Without accurate costing, valuation becomes difficult which is a risk smaller players cannot offset in the same way majors can. The 1998 Act makes no provision for this, and a Section 29 notice merely accounts for a cost estimate.

Thus, how well will the 1998 Act work when decommissioning becomes large scale? What issues arise? Davar and Shirazi present the concern of judicial review cases, arguing that with the extensive liability served under a Section 29 notice in conjunction with the lack of case law on such a fresh topic that interpreting the 1998 Act may prove cumbersome, leading to Section 29 notice holders challenging the validity of this in court.[36] Indeed, interpretation of the Act has become cumbersome as proved in the recent *Exxon v Apache*[37] case where Exxon lost their battle to require Apache to raise £100m to protect it from decommissioning liability cost. This case left questions around liability for fields with multiple phases, with new wells developed after an asset transfer. Professor Paterson questions if this case opens the risk of Section 34 not operating as a 'catch-all' protection for the taxpayer, with call-back holders able to 'point' to parts of programmes they are liable for, leaving the taxpayer liable for the rest?[38] Establishing whether a recipient falls within Section 30 may prove challenging, particularly with private equity entities which will be explored in chapter three. The possibility of this dispute may cause

---

[34] BP, 'North-West Hutton Decommissioning Programme' (*UK Government*, 7 February 2006) 20 <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/43406/nw-hutton-dp.pdf> accessed 10 February 2020.

[35] BP, 'North-West Hutton Decommissioning Programme Close-out Report' (*UK Government*, 2014) 27 <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/371545/NWH_Decommissioning_Programme_Close_Out.pdf> accessed 10 February 2020.

[36] Michael Davar and Gideon Shirazi, 'Decommissioning in the UK Continental Shelf: A Litigator's Perspective' (2015) International Energy Law Review 192, 194.

[37] *Apache UK Investment Ltd v Esso Exploration and Production UK Ltd* [2021] 4 W.L.R. 85.

[38] John Paterson, 'Apache-Exxon decom case: Is the regulator left without options? Not necessarily' (*Energy Voice,* 2021) <www.energyvoice.com/opinion/324841/apache-exxon-decommissioning-john-paterson-aberdeen-uni/> accessed 25 May 2021.

parties to be less likely to enter deals, particularly in the North Sea at its mature asset life stage. To overcome this, it is imperative for the regulatory bodies to increase scrutiny of buyer's financial position – something Australia has recently passed into law.[39]

Clean Break vs Shared

As a result of the liability arising from the 1998 Act, there has been a serious increase of consideration to the issue of exiting parties within a deal retaining decommissioning liability, on a part or whole basis. To encourage deals in the North Sea, prior to the oil price crash, sellers generally tried to achieve a clean break model with decommissioning liability, meaning the buyer accepted full liability within the Sale and Purchase Agreement (SPA) and would provide the seller with security for the costs under a Decommissioning Security Agreement (DSA). However, post oil price crash, in a low-price environment, a clean break has been difficult to achieve, driving a need for new innovative SPAs to be constructed and allow smaller entrants to acquire mature assets. Innovative SPAs have indeed been drafted to try to bypass this barrier – an example being Serica Energy acquiring interests in the Bruce, Keith and Rhum fields with BP and TotalEnergies E&P UK retaining the decommissioning liability.[40] In this deal, the bulk of the consideration and the initial cash consideration was deferred to mitigate financial risk and maintain balance sheet resilience.[41] Deals like this put less risk on smaller players which pushes deals through, thus maximising economic recovery.

It is in a multitude of parties' interests to prolong decommissioning. This is due to the fact that the removal of one installation can have a domino effect on various fields, accelerating multiple decommissioning projects. An example of this is the shut in

[39] Department of Industry, Science, Energy and Resources, *Consultation Paper: Enhancing Australia's decommissioning framework for offshore oil and gas activities'* (Australian Government 2020) 7.
[40] Serica Energy, 'Completion of the Acquisition of BP's Interests in Bruce, Keith and Rhum Fields and Re-Admission on Trading on AIM' (*Serica Energy*, 30 November 2018) <www.serica-energy.com/downloads/releases/Completion%20of%20BKR%20RNS%20FINAL.pdf> accessed 6 January 2020.
[41] ibid.

of Dunlin which meant that each of the other fields which went into the Sullom Voe terminal had increased operating costs.[42] Depending on the field, this increase in costs could have a detrimental impact to the economics of production. Does this mean that buyers will need to carry out due diligence on several linked fields and hubs before making an acquisition? This is an indirect result of decommissioning that could have an impact on deal activity within the UKCS, particularly with the high number of linked fields and terminals.

Residual Liability in Perpetuity

As briefly discussed, in addition to the large class of persons who can be trapped with initial liability, oil and gas field owners within the UKCS have a liability in perpetuity for any residues and remains which are left after decommissioning has been completed. It is clarified that it is the owner or a Section 29 notice holder who bears this liability.[43] This perpetual liability is problematic for operators within the UKCS – it is not mentioned in statute (particularly no direct mention in the 1998 Act) and the concept is only contained briefly in the Guidance Notes,[44] with the liability yet to be tested in the courts. What truly is liability in perpetuity? And does it even exist to cause a barrier and concern to parties in deals? Although it is not mentioned in legislation, BEIS Guidance continues to have significant influential power in the UK particularly in conjunction with the Secretary of State's powers to enforce. Can one really argue that a corporate entity will survive long enough to discharge this liability?

The uncertainty that emerges from this liability makes the development of practical commercial solutions to the management of UKCS operations unlikely until liability can be limited. The lines surrounding this liability are blurred – is it for carrying out what is dictated in the approved decommissioning programme? Or is it for carrying out further work that the regulator stipulates after

[42] Claire Milhench, 'Low Oil Price Domino Effect to Shut More North Sea Fields Early' (*Thomson Reuters*, 15 July 2015) <www.reuters.com/article/us-northsea-oil-decommissioning/low-oil-price-domino-effect-to-shut-more-north-sea-fields-early-idUSKCN0PP02920150715> accessed 12 October 2019.
[43] BEIS Guidance Notes (n 4) 72.
[44] ibid.

completion of the programme? If the answer is the latter, this does not seem fair and reasonable and places an uncertain and open-ended obligation upon owners even with the protection of a comparative assessment.[45] If liability in perpetuity is intended to cover future wrongs, it seems like a poor solution.

An aspect that is rarely considered is perpetual risk and liability regarding wells – plugged wells are very capable of leaking and with well work accounting for the largest proportion of decommissioning costs who does this liability and cost lie with? The introduction of s34 of the 1998 Act into the topic of liability in perpetuity means that any s29 holder could be held liable for the entire decommissioning cost or correction of problems in the future – there must be a better answer. It undeniably makes smaller players less likely to take on mature assets with the fear of this residual liability that comes with being the last owner/section 29 holder. A more sophisticated cost-effective solution would be to prepare a national or industry approach to remedying the problem in comparison to finding someone to blame and hold liable. It can constrain the transfer of assets from one operator to another more suitable one to extend the life and maintain production and thus maximise tax revenues.

It is clear that legislation is not adequate to cover the concerns regarding residual liability in perpetuity. There is little guidance and the guidance that is available is too vague. Risk is required to be diminished by establishing a clear structure and emphasis on who is responsible should something in the future go wrong.

Analysis

Decommissioning continues to be overlooked by lawmakers and there still remains an industry and regulatory wide lack of understanding regarding the best way to approach it. It is no secret that uncertainty relating to monetary policy, fiscal policy and regulation has a strong

[45] Jim Christie, Callum Falconer and Trevor Jee, 'Decommissioning Manifesto: From Here to Eternity?' (*Energy Voice*, 5 December 2017) <www.energyvoice.com/oilandgas/decom/157961/decommissioning-manifesto-eternity/> accessed 8 February 2020.

negative effect on mergers and acquisitions (M&A) activity[46] and it is imperative that clearer regulation is passed to encourage deal activity and increase investment in the North Sea. The practical, commercial, and legal uncertainties involved in offshore decommissioning makes it an area likely to be plagued with disputes.[47] Legislation must be robust, keeping in mind the possibility of early decommissioning. With the coronavirus pandemic and the collapse in OPEC+ production restraint has seen oil price reach its lowest since 2003. Instantly questions arise – with close to a quarter of fields running at a loss at the current price, early shut-ins are a real possibility, accelerating decommissioning spend by US$20bn.[48] What impact does this early decommissioning liability cause for investments? Fluctuating price environment could see annual investment in the UK fall below US$1bn as early as 2024,[49] with nearly six billion barrels left in the ground as a result of stranded assets.[50] These all act as contributing factors on the need to cease production early and trigger decommissioning.

What has the UK Government Done to Encourage Deal Activity?

Tax

Overview

The UK government have taken a multitude of steps to make the UKCS a more attractive investment hub, particularly on tax regimes, which this chapter will focus on. The taxation of production on the UKCS is notoriously complex and has been subject to frequent revision in the past 40 years.[51] This frequent changing had unfortunately earned the UK an unattractive reputation for fiscal

[46] Alice Bonaime, Huseyin Gulen and Mihai Ion, 'Does Policy Uncertainty Affect Mergers and Acquisitions?' (2018) 129 Journal of Financial Economics 531, 533
[47] Davar and Shirazi (n 36) 194.
[48] Boroujerdi (n 8).
[49] ibid.
[50] ibid.
[51] Marshall Hall, *Gas Production from the UK Continental Shelf: An Assessment of Resources, Economics and Regulatory Reform* (Oxford Institute for Energy Studies 2019) 11.

complexity and instability among upstream investors.[52] In 2014 however the government began a fiscal review[53] as the 'regime has become overly complex, and risks distorting investment'.[54] The review led to a sizeable reduction of tax in 2015 and 2016, resulting in a more competitive and attractive industry for international investors.[55] This chapter will explore the various steps by the government to create a more lucrative environment for not only investors but producers to maximise the economic recover of the remaining hydrocarbons in the basin, with a specific focus on decommissioning tax relief.

Tax Regimes

There has persistently been immense stigma from the public surrounding tax benefits that are afforded to oil and gas companies, particularly recently with the press release that Shell paid zero corporation tax in 2018[56] as a result of decommissioning tax relief, with a £60 million refund for the Brent field[57] and a total refund of $115 million.[58] Can this much tax offset be justly argued? Taking into account their $730.6 million UK profit.[59] There has to be a balance to ensure that taxpayers are not left to foot the bill, with the OGA already estimating that the public will contribute in the region of £16.89bn[60] to the total decommissioning cost of £51bn.[61] However the industry has been, and will continue to be, a significant source of revenue for the UK Exchequer, with more than £350bn being paid in the last 50 years

[52] ibid 27.
[53] HM Treasury, *Review of the Oil and Gas Fiscal Regime: Call for Evidence* (HM Treasury 2014). Also see HM Treasury, *Driving Investment: A Plan to Reform the Oil and Gas Fiscal Regime* (HM Treasury 2014).
[54] HM Treasury, *Driving Investment* (n 53) 23.
[55] ibid 58.
[56] Royal Dutch Shell, *Tax Contribution Report 2018* (Royal Dutch Shell plc 2018).
[57] Allister Thomas, 'Shell gets £60m Refund from UK Government for Brent Field Decommissioning' (*Energy Voice*, 2019) <www.energyvoice.com/oilandgas/north-sea/214225/shell-gets-60m-from-uk-government-for-brent-field-decommissioning/> accessed 18 December 2020.
[58] Royal Dutch Shell (n 56).
[59] ibid 62.
[60] Oil and Gas Authority, *UKCS Decommissioning: 2019 Cost Estimate Report* (Oil and Gas Authority 2019).
[61] ibid 64.

and more still to come.[62] Although the UKCS is a mature basin, it nevertheless contributed £1.2bn in each of the financial years 2017-18 and 2018-19, with the Office for Budget Responsibility forecasting that production tax payments by the industry will amount to more than £8.5bn over the next 5 years.[63] It must be publicly recognised that tax reliefs encourage deal activity within the UKCS with the knowledge that buyers will be free to produce at a profit as a result of decommissioning tax relief.

In a bid to increase production, Petroleum Revenue Tax (PRT) has been permanently reduced from 50% to 0% from January 2016 as the result of an amendment to Section One of the Oil Taxation Act 1975, however it has not been completely abolished in order to ensure that losses arising from PRT-liable fields could continue to be carried back against part PRT payments. Profits from a field will no longer be taxed, only when money is spent will PRT losses be activated against profits prior to 2015. This move by the government means that all UKCS fields face the same tax rate for the first time since 1982. The PRT change means that deals can be less complicated as, where the seller retains decommissioning liability, they do not have to retain an interest in the field to obtain PRT relief for the costs. This previously created an ongoing exposure to liability over and above the anticipated decommissioning liability.

Currently, oil and gas operators are subject to standard corporation tax. They are further obliged to pay Supplementary Charge Tax (SCT) which is an additional tax charged at the company level on top of Ring Fence Corporation Tax but without a deduction for finance costs. Profits which are subject to SCT may be reduced by application of the Investment Allowance or Cluster Area Allowance. SCT is however extremely difficult to understand and apply with the industry proposing that significant guidance is required. Activities on the UKCS are also subject to ring-fence corporation tax (RFCT). Ring-fence tax applies to all fields which prevents profits arising within the

[62] Oil and Gas UK, *Economic Report 2019* (Oil and Gas UK 2019).
[63] Oil and Gas UK, 'Aberdeen Can Position Itself As a Global Net Zero Energy Capital' (*Oil and Gas UK*, 5 November 2019) <https://oilandgasuk.co.uk/aberdeen-can-position-itself-as-global-net-zero-energy-capital/> accessed 30 February 2020.

ring-fence from being sheltered by losses arising from activities carried on outside the ring-fence. Expenditure incurred on decommissioning qualifies for tax relief in the form of capital allowances – the company funding the decommissioning cost must directly incur the expense. Any capital allowance claimed can be carried back and set against historic profits within the ring-fence for RFCT and SCT purposes. To do this, the entity incurring the decommissioning expenditure must have had taxable profits against which to offset these costs. This in turn made mature assets with large decommissioning liability unattractive to smaller players. What has been done to mitigate this? The government has introduced Transferable Tax History.

Transferable Tax History ('TTH')

The most prominent deal driving orientated tax change is incontestably the Transferable Tax History (TTH) which was implemented through the Finance Act 2019.[64] Schedule 15, Part 1(2)(a) states that a TTH election is for 'an amount of the seller's ring fence profits to be treated, in accordance with the provisions of this Schedule, as if it were an amount of the purchaser's profits (instead of the seller's profits)'. Highlighting that the role of TTH is to allow the company selling the field to transfer a proportion of their tax payment history to the buyer. The buyer can then set the costs of decommissioning the asset against the TTH to generate a repayment but the total TTH amount must not exceed (the lower of) the uplifted decommissioning cost estimate for the asset,[65] or the total amount of the seller's eligible ring fence profits for the period beginning 17 April 2002 and ending at the end of the reference accounting period.[66] It is activated once an asset has ceased production[67] and the total decommissioning expenditure exceeds the total net profits.[68]

The situation before TTH was introduced caused frustrations when a new entrant without a history of taxable profits acquired a

[64] Finance Act 2019, s 37 and sch 15.
[65] ibid sch 15, pt 2(4)(a).
[66] ibid sch 15, pt 2(4)(b).
[67] ibid sch 15, pt 5(30)(1)(a).
[68] ibid sch 15, pt 5(30(1)(b).

field. In this scenario, there was a risk that the decommissioning costs of the field would exceed the taxable profits generated by the new owner, thus preventing effective tax relief via the traditional carry-back mechanism, leaving the buyer in a worse position than the seller would have been in. This caused mature fields to appear unattractive to buyers and discouraged smaller and new players from producing or even entering the market; it created an uneven playing field and frustrated commercial negotiations. The aim of TTH is to extend the lives of late-life assets and delay decommissioning by encouraging investment in the UK and the UKCS.[69]

How effective is this new tax regime? The impact of TTH is heavily dependent on crude oil prices, levels of production and operating costs relative to costs of decommissioning.[70] Thus, with the recent oil price drop, the greater the impact TTH will have on a company's cash flow, it will exacerbate treasury exposure and insulate buyers of assets. It may also be argued that it does not incentivise operators to produce. This is because without TTH, a buyer has a natural incentive to invest in production as that would serve to add sufficient new taxable income to enable a carry back of decommissioning loss – without taking these actions, buyers would have to rely on high oil price which at the time of writing is unlikely. Nonetheless, Alex Kemp, a leading petroleum economist, stated that TTH is already making an impact on North Sea deals.[71] Further, Robert Jenrick, Exchequer Secretary, told the Bill Committee that the TTH mechanism 'provides new investors with the certainty that they require about the tax relief they will receive for decommissioning

[69] HMRC, 'Policy Paper: Oil and Gas Taxation: Transferable Tax History and Retention of Decommissioning Expenditure (*HMRC*, 7 November 2018) <www.gov.uk/government/publications/oil-and-gas-taxation-transferable-tax-history-and- retention-of-decommissioning-expenditure/oil-and-gas-taxation-transferable-tax-history-and-retention-of- decommissioning-expenditure> accessed 23 March 2020.

[70] Tom Mitro, 'Analysis of Transfer of Tax History (TTH) Proposal' (*Global Witness*, 28 August 2018) 9 <www.globalwitness.org/documents/19422/Transferable_Tax_History_-_Economic_analysis.pdf> accessed 15 May 2022.

[71] Alister Thomas 'New Law Making an Impact on North Sea Business' (*Energy Voice*, 2019) <www.energyvoice.com/oilandgas/north-sea/211456/new-law-making-an-impact-on-north-sea-business/> accessed 16 January 2020.

costs, to allow new deals to proceed.'[72] However others argue that it is 'fiscally irresponsible'[73] expanding the very tax breaks that put the exchequer on the hook for 'exorbitant'[74] future decommissioning liabilities. Clive Lewis argues that it favours tax cuts for the wealthiest corporations, with the taxpayer left vulnerable to huge potential pay-outs.[75] Many welcomed the new regime, stating it would allow firms to claim greater tax relief when it comes to shutting down assets, an issue which had historically been a barrier to new deals.[76] TTH neutralises CT to encourage deals where this was a barrier when bidding for mature assets with a lack of CT history. Overall, it is difficult to fully analyse the impact TTH will have on deal activity given the newness of it yet the rationale behind the passing of it is fully commendable.

Decommissioning Relief Deeds ('DRDs')

The government's introduction of decommissioning relief deeds (DRDs) have undoubtedly increased deal activity. The purpose of DRDs is to provide certainty on the tax relief that will be available to a company when decommissioning an asset. They indemnify the company for changes in tax legislation or the default joint-venture partners in respect of their decommissioning activities, allowing them to claim relief from HM Treasury. This allows companies to adopt post-tax securitisation arrangements for the future cost of decommissioning, enabling companies to transfer assets with certainty and free up significant capital for investment that would otherwise have been held in reserve for future changes in tax rules. To benefit from DRDs, a company must be a 'qualifying company' as defined in s80(3) of the Finance Act 2013. Commentators see this as a step forward for the industry, stating that the arrangement is not only to the advantage of the industry, but takes account of the overall

[72] Finance Bill (No.3) Deb 4 December 2018, col 181.
[73] ibid col 184.
[74] ibid
[75] ibid cols 187-188.
[76] Allister Thomas, 'Tax History Transfer Could Cause M&A Deal Delays' (*Energy Voice*, 2019) <www.energyvoice.com/oilandgas/north-sea/165532/tax-history-transfer-cause-ma-deal-delays/> accessed 3 December 2019.

interests of the UK.[77] The DRD will also pay out to a company that picks up the decommissioning liabilities of another company that has defaulted under its Petroleum Act obligations. Respondents to the original paper welcomed the government's objective, stating that it should help encourage and sustain investment in the UKCS, with a positive effect on transaction activity and investment in late-life assets.[78] Since its inception, at the time of writing, 92 DRDs have been agreed between the government and operators[79] and risk premiums have been reduced, unlocking £6bn of capital for reinvestment.[80] Ultimately it gives the industry much greater certainty on decommissioning liability and should facilitate a number of licence changes releasing substantial funding held under guarantee.[81]

Analysis of the Current Tax Regime

It is no secret that fiscal instability has been a significant factor in basin under-performance and a lack of transferring assets, however changes and future certainty regarding decommissioning tax relief have been well received and will help maximise long term economic recovery.[82] At a time where there is pressure to move towards clean energy, it is unlikely that the government will further ease the current fiscal terms[83] and in the 2020 budget, there was no mention of any oil and gas legislative change for the industry. With the uncertainty of the recent oil price drop, stability from the government can be seen as a plus. Derek Leith, EY's Global Oil and Gas Tax Lead said that the

---

[77] John Paterson, 'Analysis: Government Relief Deeds for Decommissioning' (*Energy Voice*, 8 April 2019) <www.energyvoice.com/oilandgas/north-sea/196553/analysis-government-relief-deeds-for-decommissioning/> accessed 22 February 2020.

[78] HM Treasury, *Decommissioning Relief Deeds: Summary of Responses* (HM Treasury 2012) 9.

[79] Oil and Gas Authority, 'Estimates of the Remaining Exchequer Cost of Decommissioning UK Upstream Oil and Gas Infrastructure' (*North Sea Transition Authority*, 2019) 3 <www.ogauthority.co.uk/media/5960/exchequer_cost_decommissioning_august_2019.pdf> accessed 20 January 2020.

[80] Oil and Gas UK, *Economic Report 2019* (n 62).

[81] Ian Wood, *UKCS Maximising Recovery Review: Final Report* (Department of Energy and Climate Change 2014) 13.

[82] ibid.

[83] Boroujerdi (n 8).

Chancellor's decision to make no change was a 'welcome move'. He stated that 'in recent years we have seen significant change to the tax landscape for the oil and gas sector, with successive governments acknowledging the maturity of the basin and the need to have stable fiscal conditions for investment'.[84]

Private Equity's Role and the Issue of Funding

Private Equity

Overview

Private equity buyers have become prominent within deals in the oil and gas industry and have given mature assets in the North Sea a new lease of life. The North Sea oil and gas business keeps evolving with new generations of finance. Why is this? Investors know that the sector has security and law, alongside a great knowledge base of debt and equity funders in the UK[85], which leads to deals and large cash injections into the UKCS. The downturn of oil and gas has been seen as a golden opportunity to buy assets which has been readily seized upon by private equity firms.[86] Private equity's interest in the sector is very apparent, with major private equity, Blackstone, raising approximately US $4.5 billion for its energy focused fund.[87] Recent deals include Neptune (owned by private equity companies Carlyle Group and CVC, and sovereign wealth fund China Investment Corporation), which approved the Seagull development in April,

[84] EY, 'Oil and gas Industry Held Steady by New Chancellor' (*EY*, 11 March 2020) <www.ey.com/en_uk/news/2020/03/oil-and-gas-industry-held-steady-by-new-chancellor> accessed 11 March 2020.

[85] Craig Guthrie, 'North Sea Asset Sales Healthy For Now' (*Petroleum Economist*, 22 September 2019) <www.petroleum-economist.com/articles/politics-economics/europe-eurasia/2019/north-sea-asset-sales-healthy-for-now> accessed 18 February 2020.

[86] Simon Flowers, 'Private Equity – Big Buyers of Oil and Gas in the Downturn' (*Wood Mackenzie*, 12 November 2019) <www.woodmac.com/news/the-edge/private-equity--big-buyers-of-oil-and-gas-in-the-downturn/> accessed 16 December 2019.

[87] Alicia McElhaney, 'Blackstone Raises $4.5 Billion for Energy Fund' (*Institutional Investor*, 14 June 2019) <www.institutionalinvestor.com/article/b1fvf6jyd10tmk/Blackstone-Raises-4-5-Billion-for-Energy-Fund> accessed 25 March 2020.

which should add another 50 000 boe/d by the end of 2021, 80% of which would be oil.[88] Another large private equity deal was the Chrysaor purchase of ConocoPhillips £2.2bn holdings[89], which was backed by Harbour private equity fund. Deals like these highlight a changing of the guard[90] in the North Sea with the ending of the dominance of oil majors such as BP, Shell and Chevron and opening the doors to smaller players. The Chrysaor deal came two years after a £3bn purchase[91] of Shell assets and has seen them rival of BP, TotalEnergies and Shell to become one of the leading producers in the North Sea.[92] With private equity firms raising more than $200 billion for energy investments since 2014,[93] it has been suggested that tax and regulatory changes discussed in chapter two have helped the latest wave of private equity deals.[94] Evidently, the changes mentioned above have driven deal activity with major equity funds but is there a major risk to the main fund that private equity have not fully evaluated?

---

[88] Mark Lamney, 'Neptune to Buy North Sea Stakes from Apache' (*Energy Voice,* 18 August 2018) <www.energyvoice.com/oilandgas/north-sea/179027/breaking-neptune-to-buy-north-sea-stakes-from-apache/> accessed 3 December 2019.

[89] Shadia Nasralla and Ron Bousso, 'Chrysaor Completes Acquisition of Conoco's UK North Sea Assets' (*Thomson Reuters*, 2019) <www.reuters.com/article/us-conocophillips-northsea-chrysaor/chrysaor-completes-acquisition-of-conocos-uk-north-sea-assets-idUSKBN1WF1JS> accessed 1 October 2019.

[90] Nevian Boroujerdi and others, 'North Sea Upstream: 5 Things to Look For in 2020' <https://my-legacy.woodmac.com/reports/upstream-oil-and-gas-north-sea-upstream-5-things-to-look-for-in-2020-369485?contentId=369485&source=21> accessed 18 December 2019.

[91] Mark Lammey, 'Chrysaor Gets £1.6bn Worth of Backing to Help Pay for Shell Assets' (*Energy Voice*, 2017) <www.energyvoice.com/oilandgas/north-sea/136414/chrysaor-lands-1-6bn-loan-help-pay-shell-assets-articleisfree/> accessed 23 February 2020.

[92] Nathalie Thomas, 'New King of the North Sea Takes on Total and BP' *Financial Times* (London, 12 October 2019) <www.ft.com/content/92e7b354-eb51-11e9-a240-3b065ef5fc55> accessed 15 May 2022.

[93] Sarah McLean, 'Private Equity Oil & Gas Transactions: Insight for Buyers and Sellers' (*Shearman & Sterling*, 21 January 2020) 4 <www.shearman.com/perspectives/2020/01/private-equity-oil-and-gas-transactions--insights-for-buyers-and-sellers> accessed 21 January 2020.

[94] Nathalie Thomas and David Sheppard, 'Private Equity Leads the Changing of the North Sea Guard' *Financial Times* (London, 13 February 2019) <www.ft.com/content/251e2496-2e0c-11e9-8744-e7016697f225> accessed 15 May 2022.

Decommissioning Liability Extending to the Fund?

Separate legal personality and limited liability are both fundamental principles of corporate law. Historically, a manager of a fund has been able to acquire and manage its subsidiary portfolio companies in the knowledge that any liabilities in the portfolio companies would (in most circumstances) be contained with no risk passing to the manager or its fund. It has however been possible for a parent company to assume liability for its subsidiary without piercing the corporate veil. Given the imperativeness for a private equity firm to keep investments separate, responsibilities for liabilities of a party among other affiliated entities is a significant concern for private equity, particularly with the cost of decommissioning.

The typical structure of a UK private equity fund is an English limited partnership governed by the Limited Partnership Act 1907. In a limited partnership there must be one or more partners with unlimited liability – these partners are called general partners who will commonly be a separate vehicle - either an English limited company or a limited partnership owned by the private equity firm.[95] While private equity fund sponsors have permanent employees, their investments and projects are not paid for by money generally held by the private equity fund sponsor itself, but rather from and for the benefit of a particular 'fund' that the private equity fund sponsor has organised by raising capital commitments from passive third-party investors. These individual funds are typically organised as limited partnerships in which the private equity fund sponsor has organised capital commitments from various limited partners who have capital available to invest in ventures. The permanent employees of the private equity fund sponsor are then positioned to invest that capital on behalf of the limited partners. The individual fund therefore does not have a bank account with case sitting to be used to fund acquisitions, it instead has capital commitments from the individual funds limited partnership agreement when money is needed for an investment opportunity.[96] Thus, due to the structure, unlike a strategic

[95] Louise Gullifer and Jennifer Payne, *Corporate Finance Law: Principles and Policy* (2nd edn, Hart Publishing 2015) 768.

[96] Erin Hopkins and Ryan Staine, 'Moneyball For The Oil Patch: How Private Equity is Re-Shaping The Upstream Oil and Gas Mergers and Acquisitions Market'

player, private equity investment is going to have an end date – private equity focus is on exiting and having the ability to distribute profits to partners and dissolve the vehicle. This is a sensitivity to private equity players in the oil and gas market. Given the nature and scale of assets involved in oil and gas deals, it is critical to have effective management and operations. Many sponsors identify a management team before they acquire an asset. Examples include CVC Capital Partners and Carlyle backing former Centrica boss Sam Laidlaw to establish the Neptune platform; Blackstone and BlueWater backing a senior team from Centrica to form Siccar Point and EIG backing Chrysaor led by Phil Kirk to acquire Shell's North Sea assets.[97] The key is to establish management teams that know how to operate mature assets in the most profitable way to prolong cessation of production.

However, Davar and Shirazi raise the concern that private equity owned oil and gas companies who are active within the UKCS fail to understand the risks that arise with decommissioning liabilities.[98] Thus, the role of private equity in the oil and gas industry is becoming more prominent, with the mix of independent and smaller private equity backed companies now active and boosting interest on the UKCS. Private equity investors aim to identify, acquire, operate, enhance, and then exit from an investment within a defined period,[99] typically 10 years, but can be as short as three or five.[100] Following this, the aim is to exit usually by way of flotation, a trade sale or sale to another private equity fund. Because of their short life span, private equity companies are usually unwilling to enter a DSA or accept any decommissioning liability under a SPA. This does not mean they are immune and the secretary of state or Section 29 notice holder may attempt to force the private equity company or entity to

(*Baker Botts LLP*, 2018) 3 <www.cailaw.org/media/files/IEL/ConferenceMaterial/2018/hopkins-paper.pdf> accessed 1 December 2019.

[97] Chrysaor, 'Chrysaor to Establish the Leading UK Independent E&P Company Focused on the North Sea: Acquisition of Asset Package from Shell for $3.0 Billion' (*Chrysaor*, 31 January 2017) <www.chrysaor.com/downloads/news/chrysaor-to-buy-north-sea-assets-final.pdf> accessed 27 March 2020.

[98] Davar and Shirazi (n 36) 197.

[99] McLean (n 93) 2.

[100] Gullifer and Payne (n 55) 770.

cover their share of the decommissioning cost. Consequently, private equity investors need to be very careful about being liable for decommissioning liability as there is the issue of potential responsibility of entities who are affiliated with the portfolio companies, including the individual funds that invested in them. Under most definitions, the individual fund would be considered an affiliate which raises the issue of liability in that the Portfolio Company's counterparty could come after the Individual Fund or even the private equity sponsor whether through a form of piercing the corporate veil argument or otherwise. There thus still remains the risk that decommissioning liability may extend to the fund itself, and not only the special purpose vehicle that was established to acquire the asset(s).[101] Measures must be put in place to mitigate the risk of exposure to reassure investors and ensure deals on the UKCS continue to happen.

In addition to decommissioning liability, residual liability (which was discussed in greater detail in chapter one) is a particular concern for private equity. It is paramount to safeguard this issue as a new generation of private equity investors with different financial priorities and opportunities other than oil and gas and will find better opportunities elsewhere.[102] Finding deal structures that reassure investors to mitigate the risk of exposure may 'hold the key'[103] to better deal activity. At the moment, insurance schemes are being explored, however companies are not keen to accept liability in perpetuity[104] and have struggled to develop a traditional insurance backed solution, mainly due to the fact that decommissioning is a certain event and not a fortuity.[105] Investors are concerned about the legal liability and the transfer of risk from the operator to the financer.

---

[101] Baker, Bartlett, Jones and others (n 5).
[102] Heffron (n 10) 189.
[103] Baker, Bartlett, Jones and others (n 5).
[104] Mark Lammey, 'Mechanism for Ending Decommissioning Liability Predicament a Priority, Top Professor Says' (*Energy Voice*, 2018) <www.energyvoice.com/oilandgas/north-sea/181035/mechanism-for-ending-decommissioning-liability-predicament-a-priority-top-professor-says/> accessed 23 February 2020.
[105] Jardine Lloyd Thompson Group, 'Decommissioning of Offshore Operations' (2019) <www.jlt.com/industry/energy-insurance/insights/decommissioning-of-offshore-operations> accessed 23 February 2020.

Questions arise surrounding residual liability in the situation where a private equity backed SPV company ceases to exist after their short investment period – will the main fund be liable? A prime risk example is well plug and abandonment. What happens in a situation where the SPV dissolves without properly and efficiently plugging and abandoning wells that were acquired from a strategic player. Strategic players may worry that the Secretary of State will come back up the chain to claim the costs. In this situation, the seller may be driven to seek ongoing credit from the buyer to support the obligations that are assumed by the buyer. This request would be difficult from a PE buyer to accommodate given their short life cycle and business model of PE backed enterprises which focus on a clean exit from investment, and any requirements that impact the PE fund sponsor's ability to exit will be difficult to accept.[106] This can frustrate deals causing less investment in UKCS. There appears to be a wave of uncertainty surrounding residual liability in perpetuity for a multitude of players and there seems to be no current solution.

Funding

Overview

The financing of decommissioning continues to be a growing apprehension for parties in deals. Inadequate provisioning of decommissioning costs could have substantial consequences for the cash flow, overall sustainability of financial position and the potential survival of the company.[107] As such, of the main challenges that will begin to arise for deals within the UKCS is the lack of access to finance,[108] not only for the deal itself but for the decommissioning liability that comes with the asset. Further, with the risk of decommissioning liability falling due to the taxpayer, there is a pressing concern around the way in which infrastructure removal is going to be funded in the next decade. There is huge uncertainty around the actual cost to taxpayers' due to the fact companies are at

[106] Hopkins and Staine (n 96).
[107] Abdo Hafez and Musa Mangena, 'Accounting Disclosures of Provisions for Decommissioning Oil and Gas Installations: The Case of Oil and Gas Companies Listed in the UK' (*Sheffield Hallam University*, 2018) 2.
[108] Flowers, *Coping With Negative Cash Flow* (n 7).

the very early stages and are still learning about the cost of different decommissioning activities.[109] Regulators take a close interest in funding for decommissioning, especially when oilfields change hands,[110] thus there must be an industry understanding of the way in which this is going to be regulated. With finance in short supply and the government reluctant to take on the burden of full cost decommissioning, leaving the industry to fully fund it[111] there must be more mechanisms in place.

Indeed, concerns in the recent case of *TAQA Bratani v Rockrose*[112] foreshadowed what happened in November 2021 when one of Brae Area partners defaulted on decommissioning liability worth millions.[113] The case highlighted operator fears around the financial capabilities of companies in deals regarding decommissioning. Although this case concerned TAQA's wish to terminate Marathon Oil's operatorship of the Brae field due to the fact RockRose (who had no operating experience) were acquiring Marathon, the passing comments in the judgement highlights the pressing funding concerns of decommissioning. Spirit Energy who was a JV partner also had 'real concerns'[114] about the financial stability of RockRose with a particular emphasis on paying the share of future decommissioning liability costs, which was a 'pressing issue'[115] for all the participants as it ran into 'many hundreds of millions'.[116] Spirit's Chief Financial Officer emphasised that other liabilities were 'less

[109] Committee of Public Accounts, P*ublic Cost of Decommissioning Oil and Gas Infrastructure* (HC 2017-19, 89) 5.
[110] Jardine Lloyd Thompson Group, 'Decommissioning of Offshore Operations' (*Jardine Lloyd Thompson Group*, 2019) <www.jlt.com/industry/energy-insurance/insights/decommissioning-of-offshore-operations> accessed 23 February 2020.
[111] Heffron (n 10) 189
[112] *TAQA Bratani v Rockrose* [2020] EWHC 58 (Comm) [105].
[113] Allister Thomas, 'Exclusive: Brae Area partner defaults on decommissioning costs after buying stake for 75pence' *The Press and Journal* (Aberdeen, 23 November 2021) <www.pressandjournal.co.uk/fp/business/3692466/exclusive-brae-area-partner-defaults-on-decommissioning-costs-after-buying-stake-for-75pence/> accessed 15 May 2022.
[114] *TAQA* (n 112) [105].
[115] ibid.
[116] ibid.

important than the risk ran on decommissioning liabilities.'[117] Consequently, the main driver of the case for transferring the operatorship to TAQA was the risk around decommissioning liability and whether RockRose would be able to meet the liability because if not, due to the joint and several nature of said liabilities, Spirit's share would increase from, 8% to 20%'.[118] This recent case highlights the mindset of operators at the moment in regards to decommissioning liability and decommissioning. It shows the complexities that can arise from acquisitions and the need for financial planning to fund the project. It shows in practice the increased risk that joint venture parties are subject to should one JV party default. The UKCS wants to remove these concerns that can act as a barrier to deal processes and perhaps put investors off with the lengthy and complex process required. It is crucial that legislation and regulation is clear enough to allow for smooth deals to complete.

What legislation exists to ensure that decommissioning is taken care of and is paid by the appropriate parties? This chapter will explore the governance of financing decommissioning and consider whether these are ample to support the industry on this unprecedented move towards large scale decommissioning.

Decommissioning Security Agreements (DSAs) – Are They Adequate?

The most prominent governance of preparing to finance decommissioning projects is through a Decommissioning Security Agreement (DSA). Decommissioning liability is a critical risk and is often addressed on an SPA or entering into a side agreement – which is the DSA. Of both these agreements, the parties agree to be liable only for an agreed percentage of the decommissioning costs, usually based on the commercial benefit that each party has derived from the asset. The structure allows each participant to pay security into a trust which will be held until the decommissioning is complete. The structure of a DSA comprises of First, Second and Third tier participants; First Tier Participants are co-ventures under a joint operating agreement and each party will pay security into the trust. Second Tier Participants are those who are at risk of falling liable to a

[117] ibid.
[118] ibid.

Section 29 notice (for example previous sellers of the asset) who remain a party to ensure that security is sufficient. Third Tier Participants are not party to a DSA, but they can enforce the terms under the Contracts (Rights of Third Parties Act) 1999.[119]

The overriding aim of a DSA is to ensure that guaranteed funds will be available at all times.[120] They have evolved as a reaction to the continuing liability created by the 1998 Act and constitutes security for the purposes of Section 38A[121] ensuring that there is a mutual protection should one or more parties fall into financial difficulties. Section 38 ensures that in the event of the insolvency of a person responsible for decommissioning, the funds set aside for meeting those liabilities remain available for decommissioning and are not accessible to the general body of creditors provided the funds have been set aside by way of security.[122] Any provision of the Insolvency Act 1986 is specifically disapplied by s38A(6) to prevent the security being used for anything other than its initial purpose of funding decommissioning. Conversely, there is no obligation to enter into a DSA[123], but there is an obligation to provide a decommissioning programme, which will require an agreement to be reached on decommissioning and will thus make provision for security. The prevailing issue with DSAs from a legislative point of view is the lack of statutory duty on operators to have financial security in place, due to the fact that the Guidance Notes are not regulation, it is mere guidance to have a DSA in place. Some second-tier operators will ensure that security is in place but not every operator is exercising this. Without a robust regime in place, s38A is ultimately useless to protect a fund if there is no fund to protect in the first place. Optimistically, there is however a requirement under section 38[124] to provide the Secretary of State information about its financial affairs and if this is not satisfactory for the duty under section 36,[125] the Secretary of State

[119] Ben Holland, 'Decommissioning in The United Kingdom Continental Shelf: Decommissioning Security Disputes' (2016) 28 Denning Law Journal 19, 22.
[120] BEIS Guidance Notes (n 4).
[121] Petroleum Act 1998.
[122] Such as a trust or other arrangement established on or after 1 December 2007.
[123] HMRC, 'Oil Taxation Manual' (*HMRC*, 13 March 2016) <www.gov.uk/hmrc-internal-manuals/oil-taxation-manual/ot28601>1 accessed 30 March 2020.
[124] Petroleum Act 1998.
[125] ibid.

may require the party to enter a DSA if this is justifiable based on concerns of financial capability.

Market conditions can cause Second Tier Participants to become concerned with the robustness of the security in place, driving them to seek higher levels of security with the apprehension that if the First Tier Participant defaults they will be left liable to cover the cost.[126] This is a forefront concern with recent market conditions as where a field may have been economically viable when the oil price was US$100 per barrel, at the recent volatile prices it may only be economically viable to produce for a shorter estimated time. Thus, previously healthy-looking security balances may now seem marginal.[127] Ultimately, this means that larger amounts of security would be required to be put into trusts each year, occurring at a time when oil and gas operators are under significant pressure from reduced revenue streams.[128]

The trepidation that BEIS have failed to adequately monitor the methodology underpinning the amounts paid by participants under a DSA has been raised with the argument that this has led to a historic under-provision in decommissioning security.[129] Requesting larger amounts of security to be retained is locking away capital that could be utilised in other areas, such as energy transition projects to ensure the UK is complying with the Paris Agreement to reach net zero. How can legislation fix this issue? Can it in truly fix it in such an unpredictable industry? The stability and predictability of the fiscal regime in the UK are core components of the competitiveness and attractiveness as an investment proposition in this global industry.[130] Without certainty on who will cover the bill, this will slow down deals and assets changing hands, particularly to smaller players, with sellers carrying the angst that the Secretary of State liability will come back to them. There is also the public concern to address that the

[126] Holland (n 119) 19-36.
[127] ibid.
[128] ibid.
[129] Greg Gordon, John Paterson and Emre Üşenmez, *UK Oil and Gas Law: Current Practice and Emerging Trends* (3rd edn, Edinburgh University Press 2018).
[130] HM Treasury, *Driving Investment* (n 53).

government needs to ensure operators and their partners have relevant finance to avoid taxpayers being left with the bill.[131]

At the best of times, quantification of decommissioning is complicated and tentative with the risk of uncertain financials accounted for resulting in an unnecessary excessive capital being retained, or on the other side of the scale, projects running over cost and leaving operators with millions less security than accounted for. There is a pressing need to establish a regime that is robust and removes uncertainty for investors in both the UKCS and technology, whilst also creating certainty for the government. Residual liability was discussed in chapter one – is it fair and reasonable to impose such an elongated time liability on owners and should they be required to financially prepare for the possibility of this?

Can one truly say that the current regimes in place have succeeded to adequately prepare for the world's largest decommissioning period? Apache recently required a £530m loan to pay off decommissioning obligations in the North Sea after Standard and Poor's downgraded the operator's credit rating from BBB to BB+.[132] This leaves the question – what security in trust under DSAs do Apache have if a £530m loan is required? This effectively shows the lack of impact DSAs are having within the industry as a mechanism to ensure that operators are financially prepared.

Thus, can it really be said that DSAs do the job? Irrefutably, the structure and thesis behind them is successful and leaves very little to fault, however could more regimes be put in place to enforce compliance? Absolutely.

---

[131] Li Jia, Peng Yun, Zhu Ming, and others, 'Decommissioning In Petroleum Industry: Current Status, Future Trends and Policy Advice' (2019) 237 Earth and Environmental Science.

[132] Allister Thomas, 'Apache Calls In £530m Loan to Pay For North Sea Decommissioning After S&P Downgrades Credit Rating' (*Energy Voice*, 27 March 2020) <www.energyvoice.com/oilandgas/north-sea/231325/apache-calls-in-530million-loan-to-pay-for-north-sea-decommissioning-after-sp-downgrades-credit-rating/> accessed 27 March 2020.

Analysis

It is evident that funding decommissioning and the risk that the Petroleum Act 1998 poses to all participants is prevalent. Although international organisations influence a high proportion of domestic law on the technicalities of decommissioning, the financial framework is surprisingly untouched from the international level. This puts the UK in a great position to hold operators accountable. Indeed, it is seen through the 1998 Act that the Secretary of State has an extensive power to bring in a wide-ranging list of bodies who are financially liable. DSAs ought to be governed better, however it can also be argued that the whole financial framework is something that could perhaps be agreed on at the international level, allowing each jurisdiction to follow a similar payment regime, closing this gap in the industry that has yet to be addressed.

Conclusion

To conclude, it has been made apparent throughout this paper that decommissioning is a tremendously topical subject area and there needs to be an immense focus given to the law surrounding it and industry practice in the near future. Based on the aforementioned deals (and many uncommented deals), it would be unjust to conclude that decommissioning liability is an absolute barrier to deal activity within the UKCS, however it also cannot be affirmed that it is not a concern for investors. With this in mind, there ought to be more research and regulation put in place to govern this area to restore confidence in investors to make the UKCS a more attractive business venture.

The steps taken by the UK government to implement new and change existing tax regimes has undeniably made the UKCS more attractive to buyers – specifically smaller players and private equity backed firms. Allowing a tax claim resulting from decommissioning has given the North Sea a new life and allowed the economics of a field to live longer than it would have without such tax breaks.

The biggest issue with decommissioning liability is the financial aspect. It has been noted throughout this paper that there has

been a severe lack of preparation for decommissioning to ensue within the industry. Available finance and cash flow is sparse and all operators (and entities) that have been associated with an asset are extremely vulnerable to liability from the secretary of state. How long will it be until multiple decommissioning projects get delayed, causing more expense to operators, and exposing the taxpayer to even more risk. Mature assets will not sell if it is not economically viable to produce for the buyer and the sellers are also concerned that the secretary of state can come back to them with a Section 29 notice. Consequently, legislation and regulators have not implemented robust mechanisms to prepare for the financing of decommissioning oil and gas installations. This has become particularly apparent in the recent oil crisis with security in trusts falling short. Accordingly, this paper has explored and recommended that the regulators put more statutory duties on operators to prepare and have security in place at the earliest possible time for each asset producing, similar to that of the nuclear industry. Legislation should be tightened to make DSAs compulsory, coupled with greater scrutiny by the OGA on the available funds each operator has in trust. The risk may even arise that any 'decommissioning fund' within an operator which is in place will be used in pressing times like the current economic climate to fund other areas, such as production. A strengthening of legislation would prevent this from happening.

Overall, although decommissioning liability is a concern for asset buyers, it certainly has not 'seen the death of the North Sea'.[133] Changes to the current regime must 'balance the desire to drive better performance and recovery with the risk of discouraging investment'.[134] It must be ensured that the UKCS 'remains an attractive destination for investment' whilst taking care 'not to impose any unnecessary additional bureaucracy'.[135]

[133] Guthrie (n 85).
[134] Wood (n 81) 18.
[135] ibid.

Coronavirus and Corporate Insolvency: A Comparative Analysis of the Corporate Insolvency and Governance Act 2020

NIGEL J ISHERWOOD*

Abstract

Since the 1980s, the emphasis of the UK insolvency regime has been on taking early steps to confront and remedy corporate troubles. The Corporate Insolvency and Governance Act 2020, enacted partly in response to the COVID-19 pandemic, represents the greatest change to the UK insolvency regime in over 20 years. Early case law demonstrates the utility of some of these changes. The pandemic response in the UK has been similar to that of Germany. Both differ greatly from the response in the US, which has adopted minimal legislative change. In the short term, it is anticipated there will be a high number of companies interacting with the UK insolvency regime, while the long-term direction is unknown due to the historic trend of stagnated development and recent developments ostensibly reactionary to COVID-19.

Keywords: corporate insolvency, business restructuring, COVID-19, United Kingdom, Germany, USA

Introduction[1]

The United Kingdom, like many other established economies, is debt based. Society enables companies to access credit facilities, in doing so

* LLB with English law (Hons) (First Class) University of Aberdeen, 2021. The author is currently undertaking the Bar Course at The Inns of Court College of Advocacy.

[1] The law as discussed is correct at the time of writing in April 2021. There has almost inevitably been further case law relating to the measures introduce by the Corporate Insolvency and Governance Act 2020. Cases after this date have not been overlooked but rather came after the fact of writing. A note of subsequent cases can be found at the end of the article.

exposing lenders to the risk of a company's future inability to pay its debts as they fall due.[2] The UK insolvency regime[3] is tasked with a multitude of objectives.[4] The primary objective of the regime is to replace the free-for-all race to enforce by creditors. The regime comprises a robust set of rules in which creditors' rights and remedies are suspended and the debtor's assets are collected, realised and distributed in an orderly fashion according to the ranking of the creditors' claims.[5] This article will be structured as follows. Part one will analyse the origin of the UK insolvency regime, identifying the Cork Report and subsequent Insolvency Act 1986, the Enterprise Act 2002 and Companies Act 2006 as the key developments. Part two will identify the changes to the regime brought about during and as a result of COVID-19 and analyse the cases of *Re Virgin Atlantic Airways Ltd*[6] and *Re Deepocean 1 UK Ltd*.[7] Part three examines the circumstances surrounding the case of *Re Virgin Atlantic Airways*[8] and compares the response of the UK discussed in part two to that of the United States and Germany. It is identified that the response of the UK has been replicated by Germany and rejected by the US. The US and Germany were chosen because they cover a broad spectrum of approaches to debtor and creditor orientated insolvency procedures, and due to the distinct interaction both regimes have with that of the UK. Part four analyses the anticipated short-term spike in companies engaging with the UK insolvency regime as support measures are eased, with the long-term direction of the regime, largely unknown.

---

[2] Kenneth Cork, *Report of the Review Committee on Insolvency Law and Practice* (Cmnd 855, 1982) ch 1.

[3] Insolvency laws are partially devolved to Scotland, with different rules in respect of the devolved elements. This article talks about the UK regime generally without focus on the devolved insolvency rules in Scotland.

[4] Ewan McKendrick (ed), *Goode on Commercial law,* (5th edn, Penguin 2017) 876-881; Vanessa Finch and David Milman, *Corporate Insolvency Law: Perspectives and Principles*, (3rd edn, CUP 2018) 25

[5] Finch and Milman (n 4) 1.

[6] *Re Virgin Atlantic Airways Ltd* [2020] EWHC 2191 (Ch), *Re Virgin Atlantic Airways Ltd* [2020] EWHC 2376 (Ch).

[7] *Re Deepocean 1 UK Ltd* [2021] EWHC 138 (Ch).

[8] *Re Virgin Atlantic Airways Ltd* (n 6).

The UK Insolvency Regime Pre COVID-19

Background

The origins of UK corporate insolvency law can be traced to the nineteenth century and the development of companies having a distinct legal identity.[9] The Limited Liability Act 1855 established the concept of company members having limited liability for debts incurred by the company. This was confirmed by the House of Lords in *Salomon*,[10] where it was held that the legal personality of a duly formed company was a distinct legal person separate from its members.

During the early half of the twentieth century, a confused tangle of insolvency laws existed that were prone to manipulation and difficult to operate.[11] Various committees were established to tackle this with limited success.[12] It was not until the mid-1970's that the deficiencies in insolvency law were attended to.[13] This was largely due to the UK's accession to the European Economic Community (EEC),[14] which necessitated the UK negotiate with other Member States concerning the drafting of the EEC Bankruptcy Convention. An advisory committee for the Department of Trade was appointed in 1973 under the chairmanship of Sir Kenneth Cork. The advisory committee's report found that a comprehensive review of the UK's insolvency framework was required, not only to participate in EEC negotiations, but because the state of the law demanded it.[15] In January 1977, the Cork Committee was asked to review, examine and make recommendations on: the law and practice relating to

[9] The Joint Stock Companies Act 1844 established the 'company' as a distinct legal identity, although it retained unlimited liability for the shareholders. Corporate insolvency was dealt with by means of special statutory provision in the Companies Winding Up Act 1844.
[10] *Salomon v A.Salomon & Co Ltd* [1897] AC 22.
[11] Finch and Milman (n 4) 12.
[12] Crowther Report (Cmnd 4596, 1971); Payne Report (Cmnd 3909, 1969).
[13] In 1975 'Justice' published a report highlighting a number of serious deficiencies in the law and made a number of reform proposals, some of which were adopted in the subsequent Insolvency Act 1986: Justice, *Bankruptcy* (Stevens & Sons 1975).
[14] The European Communities Act 1972.
[15] Report of the Cork Advisory Committee (Cmnd 6602, 1976).

insolvency; bankruptcy; liquidation; and receivership, with the view of formulating a comprehensive insolvency system.[16]

The Cork Report

The Cork Report[17] was published in its final form in June 1982.[18] The report proposed fundamental reforms central to which was a unified insolvency code to replace the array of existing legislation. Cork also proposed reforms designed to increase the survival chances of companies in financial difficulties, believing more companies would be saved if outside administrators could be brought in before a bank would formally appoint a receiver, and in circumstances when the company lacked a loan structure allowing the appointment of a receiver.[19] The Cork Report also introduced the concept of an 'administrator' to corporate insolvency with the function of managing a company's business during a period of grace (moratorium) in the hope of restoring the company to profitability.

The Cork Report provided the most comprehensive review of the UK insolvency regime ever undertaken. The central premise was a system that would facilitate corporate rescue rather than process corporate failure.[20] Cork's philosophy ultimately represented an increased emphasis on the rehabilitation of financially distressed companies and the inception of a UK 'rescue culture'. In 1986 the Insolvency Act (IA) implemented the unified code of insolvency law

---

[16] Cork, *Report of the Review Committee* (n 2) 3, On the background to, and implementation of Cork see Bruce Carruthers and Terence Halliday, *Rescuing Business: The Making of Corporate Bankruptcy Law in England and the United* (OUP 1998) 112-123.

[17] Cork, *Report of the Review Committee* (n 2).

[18] In 1979 the Cork Committee issued an interim report to the Minister, published in July 1980 as, *Bankruptcy: interim Report of the Insolvency Law review Committee* (Cmnd 1968, 1980). The Government also produced a Green Paper: *Bankruptcy: A Consultative Document* (Cmnd 9768, 1980). The Green Paper contained proposals for the privatisation of insolvency procedures which were attacked by commentators (see Ian Fletcher, 'Bankruptcy Law Reform: The Interim Report of the Cork Committee, and the Department of Trade Green Paper' (1981) 44 Mod L Rev 77) and subsequently dropped.

[19] Kenneth Cork, *Cork on Cork: Sir Kenneth Cork Takes Stock* (Macmillan 1988) 184-203.

[20] Cork, *Report of the Review Committee* (n 2).

advocated by Cork. The IA is the principal insolvency statute, save for provisions regarding the disqualification of directors,[21] and schemes of arrangement.[22] Despite the recommendations of the Cork Report not being implemented *verbatim*, the IA has endured several economic recessions and been subject to numerous reviews,[23] amendments[24] and supplements.[25]

Millennial Developments

The Enterprise Act 2002 (EA) implemented a number of highly significant changes to the UK insolvency regime, the flagship being the phasing out of administrative receivership to be replaced entirely by administration.[26] Under the administration process, insolvency practitioners are obliged to act in the interest of the creditors of the company as a whole, rather than in pursuit of the floating charge

---

[21] Company Directors' Disqualification Act 1986. A few provisions of the Companies Act 2006 are relevant to insolvency and survive the Insolvency Act 1986: Companies Act 2006, ss 754, 895-901L and 993.

[22] Companies Act 2006, pts 26 and 26A.

[23] See *Company Voluntary Arrangements and Administration Orders: A Consultative Document* (Insolvency Service 1993);
Revised Proposals for a New Company Voluntary Arrangements Procedure (Insolvency Service 1995);
A Review of Company Rescue and Business Reconstruction Mechanisms: Report by the Review Group (Insolvency Service 2000);
Department of Trade and Industry, *Productivity and Enterprise: Insolvency – A second chance* (Cm 5234, 2001); and
Company Law Review Steering Group, *Modern Company Law for a Competitive Economy: Final Report* (Department of Trade and Industry 2001).

[24] Key amending legislation since 1986 has included the Insolvency Act 1994, the Insolvency Act (No 2) 1994, the Insolvency Act 2000, the Enterprise Act 2002, the Companies Act 2006, the Enterprise and Regulatory Reform Act 2013, the Deregulation Act 2015, the Small Business, Enterprise, and Employment Act 2015 and the Corporate Insolvency and Governance Act 2020.

[25] Company Directors Disqualification Act 1986, The Insolvency Rules 1986 (SI 1986/1925),
The Insolvency (England and Wales) Rules 2016 (as amended) (SI 2016/1024), Bankruptcy (Scotland) Act 2016, The Insolvency (Scotland) (Receivership and Winding up) Rules 2018 (SI 2018/347), The Insolvency (Scotland) (Company Voluntary Arrangements and Administration) Rules 2018 (SI 2018/1082); in addition to a plethora of subsidiary statutory instruments.

[26] With limited exceptions. See, Enterprise Act 2002, s 250 inserting ss 72A-72G into the Insolvency Act 1986.

holders' interest alone. Additionally, the EA provided a 'prescribed part' for unsecured creditors[27] and ended the Crown's status as a preferential creditor.[28] The Crown's preferential status has however, since been partially re-introduced in the Finance Act 2020.[29] The EA promoted UK rescue culture by encouraging those involved with potentially financially distressed companies to confront insolvency risks in advance of the final crisis; to manage risks *ex-ante* rather than *ex-post.*[30] It follows that there is now an increased demand to audit and risk manage more generally across public and private sector activities.[31] The EA represented a shift away from debt collecting,

---

[27] On the 'new' administration procedure see Finch and Milman (n 4) ch 9;
Ian Fletcher, 'UK Corporate rescue: Recent Developments Changes to Administrative Receivership, Administration and Company Voluntary Agreements the Insolvency Act 2000, the White Paper 2001 and the Enterprise Act 2002' (2004) 5 EBOR 1999; and
Vanessa Finch, 'Re-invigorating Corporate Rescue' [2003] JBL 527;
Vanessa Finch, 'Control and Coordination in Corporate Rescue' (2005) 25 Legal Stud 374.
On the same procedure as a route to winding up see Andrew Keay, 'What Future for Liquidation in the Light of the Enterprise Act Reforms?' [2005] JB 143; and
Lisa Linklater, 'New style Administration: A substitute for liquidation? (2005) 26 Co Law 129.
On reforms dealing with administrative receivership and the ring-fenced fund see Insolvency Act 1986, ss 72A-G and 176A and sch B1; and
Insolvency Act 1986 (Prescribed Part) Order 2003 (SI 2003/2097).

[28] Enterprise Act 2002, s 251 (Insolvency Act 1986, sch 6, paras 1, 2, 3-5C, 6, 7 are deleted). For discussion see Finch and Milman (n 4) ch 14.

[29] Crown preference was partially re-introduced and came into effect on 1 December 2020. HMRC has the preferential right to be paid ahead of floating charge holders and unsecured credits, payments due to VAT, PAYE and NI contributions. For discussion see chapter two, below.

[30] On the rise of pre-packaged administrations and the 'pre-pack' as an aspect of the movement towards anticipatory action see Finch and Milman (n 4) ch 10;
Vanessa Finch, 'Pre-Packaged Administrations: Bargains in the Shadows of Insolvency or Shadowy Bargain? [2006] JBL 568;
Vanessa Finch, 'Pre-packaged Administrations and the Construction of Propriety' [2011] JCLA 1; and
Vanessa Finch, 'Corporate Rescue: A Game of Three halves' (2012) 32 Legal Stud 302.

[31] Turnbull Report, *Internal Control Guidance for Directors on the Combined Code* (ICAEW 1999); and
The Corporate Governance Committee, *The Report of the Cadbury Committee on The Financial Aspects of Corporate Governance* (Gee 1992).

towards the management and controlling of insolvency risks,[32] with commentators arguing that corporate rescue demands external intervention be engaged at the earliest opportunity.[33]

The Cork Report stated that a functional modern system of insolvency law should provide a means of rescuing viable commercial enterprises capable of making a useful contribution to the economy of the country.[34] The process of rescue describes 'a major intervention necessary to avert eventual failure of the company'.[35] Rescue has been endorsed by bankers,[36] politicians and the judiciary[37] in cases such as

[32] Corporate Governance Committee (n 31) also demand that boards identify risks to the company's value and state how these are managed. See also John Parkinson 'Disclosure and Corporate Social and Environmental Performance. Competitiveness in a Broader Social Framework' (2003) 3 JCLS 3, 6-11.

[33] Cork, *Report of the Review Committee* (n 2) ch 9; and
Cork, *Cork on Cork* (n 19) ch 10.

[34] Cork, *Report of the Review Committee* (n 2).

[35] Alice Belcher, *Corporate Rescue* (Sweet & Maxwell 1997)12;
Alice Belcher, 'The Economic Implications of Attempting to Rescue Companies' in Harry Rajak (ed), *Insolvency Law: Theory and Practice* (Sweet & Maxwell 1993);
Muir Hunter, 'The Nature and Functions of a Rescue Culture' (1999) 104 Com LJ 426;
Vanessa Finch, 'Corporate Rescue Process: The Search for Quality and the Capacity to Resolve' [2010] JBL 502;
Vanessa Finch, 'Corporate Rescue: who is Interested?' [2012] JNL 190; and
Sandra Fisby, 'Insolvency Law and Insolvency Practice and Pragmatism Diverge?' (2011) 64 CLP 349.

[36] The British Bankers' Association publicly endorsed a rescue culture. *Banks and Businesses Working Together* (British Bankers' Association 1997) para 3: 'banks have long supported a rescue culture and thousands of customers are in business today because of the support of their banks through difficult times'.

[37] *Powdrill and Another v Watson and Another* [1995] 2 AC 394 (Lord Browne-Wilkinson).
On the development of the rescue culture see, Insolvency Service, *A Review of Company Rescue and Business Reconstruction Mechanisms,* (Department for Trade and Industry 2000) 12-23,
For further judicial reference to the rescue culture, see, *Re Demaglass holdings Ltd* [2001] 2 BCLC 633 (Neuberger J); and
On Demand Information plc (in administrative receivership) and another v Michael Gerson (Finance) plc and another [2000] 4 ALL ER 734 (Robert Walter LJ)

Spectrum Plus[38] and *Leyland Daf*.[39] These cases served to either incentivise different financing arrangements or to promote a legislative response. However, the success of rescue is subjective; a company's directors, creditors, shareholders, employees, suppliers, customers and tax authorities all have their own distinct visions of a successful rescue.[40] The Insolvency Services 2016 'Review of Corporate Insolvency Framework',[41] as well as the stated aims of the EA,[42] all relate to the preservation and rescue of companies in financial difficulty. Yet there is no standardised result of the rescue process.[43] A company may be restored to its former state,[44] but it is more likely

[38] *National Westminster Bank plc v Spectrum Plus Limited and others and* others [2005] UKHL 41.

[39] *Buchler and another (as joint liquidators of Leyland Daf Ltd) v Talbot and another (as joint administrative receivers of Leyland Daf Ltd) and others* [2004] UKHL 9.

[40] Ultimately rescue is not so much about laws and processes as the cultures prevalent within the body of creditors. See Sandra Frisby, *Report to the Insolvency Service: Insolvency Outcomes* (Insolvency Service 2006); and
Stuart Slater, David Lovett and Laura Barlow, *Leading Corporate Turnaround* (John Wiley & Sons Ltd 2006).

[41] *A review of the Corporate Insolvency Framework* (Insolvency Service, London 2016). For a summary of the responses to the Consultation, see *Summary of Responses: A Review of the Corporate Insolvency Framework* (Insolvency Service 2016). For comment, see Christopher Umfreville, 'A review of the Corporate Insolvency Framework: A New Moratorium to Help Business Rescue?' (2016) 385 Sweet and Maxwell's Company Law Newsletter 1;
Tom Astle 'More Questions than Answers: The consultation on the Corporate Insolvency Framework [2016] 9 Corporate Rescue and Insolvency 133; and
John Tribe, 'The Not so Great Reformation?' [2016] Recovery (Autumn) 30.

[42] House of Commons Business and Enterprise Committee, *The Insolvency Service: Sixth Report of Session 2008-9* (HC 198, 6 May 2009) 11.

[43] Belcher, *Corporate Rescue* (n 35) 24-34, Davis Brown, *Corporate Rescue*, (John Wiley & Sons 1996) 6-8.

[44] On the market for corporate control see, Julian Franks and Colin Mayer, 'Capital Markets and Corporate Control: A Study of France, Germany and the UK' (1990) 10 Economic Policy 191-231; and
Caroline Bradley, 'Corporate Control: Markets and Rules' (1990) 53 MLR 170.

to be reorganised,[45] restructured,[46] downsized,[47] subjected to sell-offs,[48] or take-over.[49]

Commercially and politically, there has been a consolidation of rescue culture in the UK, and a new emphasis on managing insolvency risks proactively. This has led to pre-packaged administrations rapidly growing in popularity.[50] However, some commentators have critiqued that too little attention is paid to the damage to supplier companies and the wider economy in the creation of 'zombie' companies through pre-packaged administrations.[51]

Informal Rescue

In terms of promoting a rescue culture, these programmes of legislative reform cannot be judged as an unqualified success. In the aftermath of the EA, there has been a decline in the use of formal rescue procedures such as Company Voluntary Agreements (CVAs) and administration and an increase in informal rescues.[52] Informal rescue demands that parties with contractual rights agree to compromise, waive, defer debts, or alter priorities. Accordingly, dissenting creditors have the power to thwart informal rescue by triggering formal insolvency procedures. This renders informal rescue a fragile device dependent on a high degree of co-operation from a

---

[45] Where, for example, managerial reforms are instituted.
[46] Where, for example, closures of elements of the business are involved.
[47] Where operations may be cut back, workforces reduced or rationalised.
[48] Where parts of the business are sold to other firms or even to managers in management buyouts (MBOs).
[49] Where the market for corporate control operates with regard to a troubled company and a takeover prompts drastic managerial change.
[50] 'Pre-pack sales in administration report and draft regulation' (T*he Insolvency Service*, 8 October 2020) <www.gov.uk/government/publications/pre-pack-sales-in-administration> accessed 4 April 2021.
[51] Finch (n 4) 117: 'Zombie' companies are businesses that can service its interest obligations but are unlikely ever to be able to repay their principal debt; or a company whose pension scheme deficits are greater than the cashflow ever likely to be generated; or a company whose existing debts negate its ability to attract further capital.
See also: Businesses and Enterprise Committee, *The Insolvency Service: Sixth Report of Session 2008-9* (HC 2008-09, 198) 11.
[52] Sandra Frisby, R*eport to the Insolvency Service: Insolvency Outcomes* (Insolvency Service 2006).

range of parties. It is almost impossible to separate out informal rescue activity from routine negotiations.

A popular informal rescue option is the 'London Approach' which co-ordinates creditors' agreements in accordance with informal guidelines. The Bank of England acts as an 'honest broker' in making efforts to persuade reluctant parties to agree to informal settlements.[53] The London Approach has been said to have four main tenets:[54] the banks are supportive and do not rush to appoint receivers; information is shared amongst all parties to the rescue; banks and other creditors work in a co-ordinated fashion to reach a collective view; and pain is shared on an equal basis. The value of the London Approach has, however, been largely confined to very large rescue attempts with extensive borrowings.[55]

Formal Processes

The UK marketplace has responded to the promotion of rescue and recovery by lenders and the government[56] by providing services that are designed to prevent corporate failure. This has taken the form of a variety of new specialists:[57] turnaround professionals, company doctors, consultants, solutions providers, independent business reviews, asset-based lenders, private equity and others.[58] The formal arrangements available to companies in financial distress are provided for in the Insolvency Act 1986[59] and Companies Act 2006.[60]

---

[53] Finch and Milman (n 4) ch 7, In 1998 the Financial Services Authority took over from the Bank of England as banking regulator. With the breaking up of the FSA in 2012, the Prudential Regulation Authority now undertakes this role.
[54] Belcher, *Corporate Rescue* (n 35) 118.
[55] Robert Haugen and Lemma Senbet, 'Bankruptcy and Agency Costs' (1988) 23 Journal of Financial and Quantitative Analysis' 27.
[56] The Insolvency Service, *Productivity and Enterprise: insolvency – A second chance* (Cm 5234, 2001).
[57] Vanessa Finch, 'Doctoring in the Shadows of insolvency' [2005] Nov JBL 690.
[58] Ian Gray, 'The Doctor will see you now' [2014] Aut Recovery 29.
[59] Insolvency Act 1986, pts A1-XIX; Finch and Milman (n 4) chs 9 and 11.
[60] Companies Act 2006, s 895.

Administrative Receivership[61]

Prior to the EA, creditors who had lent money to a company and secured it by means of a floating charge over the whole, or substantial whole of the company's assets[62] could appoint an administrative receiver. The administrative receiver's primary duty was to his appointer, to realise their security.[63] After deducting his remuneration, expenses and paying prior ranking creditors, he would pay the proceeds to his appointor up to the amount of the secured debt and pay any balance to subsequent ranking creditors, the company or its liquidator. The EA largely replaced receivership with administration. Official figures from the Insolvency service indicate that there were 12 administrative receiverships between 2016 and 2020.[64]

Administration[65]

A court-based procedure introduced by the Insolvency Act 1985 following the Cork Committee's identification of the benefits that could flow from having a corporate insolvency procedure designed specifically for corporate rescue, rather than asset realisation.[66] The revisions made by the EA streamlined the process[67] in which a company can be put into administration by the court.[68] The court must be satisfied that the company is, or is likely to be unable to pay its debts before making an order to appoint an administrator, except if the application is from the holder of a qualifying floating charge.

---

[61] Insolvency Act 1986, ss 18-69, 72H.
[62] ibid s 19(2).
[63] Finch and Milman (n 4) ch 3.
[64] *Quarterly Company Insolvency Statistics, Q4 October to December 2020* (The Insolvency Service 2021).
[65] Insolvency Act 1986, sch B1. See also, Gavin Lightman, Gabriel Moss and Ian Fletcher, *The Law of Administration and Recievers of Companies* (5th edn, Sweet & Maxwell 2011); and McKendrick (n 4) ch 11.
[66] Cork, *Report of the Review Committee* (n 2) paras 29-33, chs 6 and 9.
[67] McKendrick (n 4) 895.
[68] On application by the company, its directors or one or more creditors, out of court by a holder of a qualifying floating charge, or out of court by the company or its directors.

Winding-Up and Liquidation

Liquidation is a procedure of last resort. Once the process has been completed, the company is dissolved. A liquidator has no powers to carry on the company's business except for the purpose of winding up.[69] Liquidation can be compulsory or at the instance of creditors.[70] Compulsory liquidation is by order of the court and the only method by which a creditor can initiate winding up.[71] Creditors' voluntary liquidation requires a resolution by shareholders to put the company into liquidation followed by a creditor decision to appoint a liquidator, and establish a liquidation committee whose members are principally creditors representatives.

Formal Arrangements with Creditors

Such agreements may defer payments or postpone collection (a moratorium); they may agree to pay sums less than those due (a composition) or pay a designated sum where there is doubt about the quantum or enforceability of a claim (a compromise). Typically, the settlements contain a variety of terms and operate within a statutory or contractual format between the company and its creditors.[72] Formal statutory arrangements can be made under the Companies Act 2006 (CA) Part 26 (or now Part 26A), section 895, while CVAs are made under IA Part 1 section 1.[73]

Part Conclusion

The UK insolvency regime has become highly sophisticated with companies able to make use of formal and informal rescue processes,

[69] Insolvency Act 1986, sch 4, para 5.
[70] Finch and Milman (n 4) ch 13.
[71] A winding-up petition can be presented by a creditor, the directors, the company's shareholders and, in certain circumstances the Departments for Business, Energy and Industrial Strategy or Financial Conduct Authority in the public interest, for example, to stop enterprises trading where they engaged in practices that defraud customers and swindle the vulnerable.
[72] Finch and Milman (n 4) ch 7.
[73] Insolvency Act 1986, ss 1-7.

and ultimately the statutory processes of winding up.[74] Recent emphasis has been on taking early steps to confront corporate troubles and to effect rescues and turnarounds before there is any need for formal action. Overall, the central purpose of the UK insolvency regime is to preserve the value of a company where that is likely to be greater than the liquidation value.

UK response to COVID-19

Context

On 11 March 2020, the World Health Organisation declared the outbreak of COVID-19 to be a pandemic.[75] Twelve days later, the UK went into lockdown.[76] On 28 March 2020, the UK government announced that it would be introducing legislative changes to insolvency laws in response to COVID-19.[77] The Corporate Insolvency and Governance Act 2020 (CIGA) came into force on 26 June 2020. The CIGA represented the greatest change to the UK's corporate insolvency regime in more than 20 years.[78] Contained within CIGA were several temporary provisions to assist companies and their directors during the COVID-19 pandemic, and a number of permanent provisions designed to significantly bolster the UK's restructuring arsenal. Amongst other things CIGA implemented measures contained in the UK government's 2018 consultation on

---

[74] Pontain Okoli, Rescue Culture in the United Kingdom: Realities and the need for a Delicate Balancing Act (2012) 23(2) International Company and Commercial Law Review 64.

[75] 'Coronavirus Disease (COVID-19) events as they happen' (*World Health Organisation*, 31 July 2020) <www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen> accessed 10 February 2021.

[76] 'Prime Minister's statement on coronavirus (COVID-19) 23 March 2020' (*Prime Minister's Office*, 23 March 2020) <www.gov.uk/government/speeches/pm-address-to-the-nation-on-coronavirus-23-march-2020> accessed 14 February 2021.

[77] 'Government introduces legislation to relieve burden on businesses and support economic recovery' (*Department for Business, Energy and Industrial Strategy* (BEIS), 20 May 2020) <www.gov.uk/government/news/government-introduces-legislation-to-relieve-burden-on-businesses-and-support-economic-recovery> accessed 14 February 2021.

[78] 'Major changes to insolvency law come into force' (*The Insolvency Service*, 29 June 2020) <www.gov.uk/government/news/major-changes-to-insolvency-law-come-into-force> accessed 24 March 2021.

'Insolvency and Governance'[79] and the Insolvency Service 2016 'Review of Corporate Insolvency Framework'.[80]

Temporary Measures

CIGA temporarily restricted winding-up petitions, suspended director liability for wrongful trading and enabled companies to adopt a flexible approach to general business formalities.[81] During the period 1 March 2020 through 30 September 2020,[82] and 26 November 2020 through 30 April 2021,[83] wrongful trading provisions were suspended and the courts were required to assume that a director is not responsible for any worsening of the financial position of the company or creditors.[84] This suspension reduced, but did not remove, the risk of directorial personal liability and applied only to the directors of certain types of companies[85] in relation to actions taken during the suspension period. The suspension was intended to prevent directors from undertaking insolvency proceedings prematurely out of concern personal liability might otherwise arise if the company fell into insolvency and had continued trading during the pandemic. However, CIGA has not suspended other areas of liability relevant to a director's conduct including fraudulent trading,[86] transactions defrauding creditors,[87] misfeasance,[88] or the general duties of a director.[89] Even so, it was possible for directors of companies with no reasonable prospect of avoiding insolvency

[79] *Insolvency and Corporate Governance* (BEIS 2018); and *Insolvency and Corporate Governance: Government response* (BEIS 2018).
[80] Insolvency Service, *A review of the Corporate Insolvency Framework* (n 41).
[81] The requirement of holding shareholders' meetings and extension of time to make certain filings at Companies House was extended by The Corporate Insolvency and Governance Act 2020 (Coronavirus) (Extension of the Relevant Period) Regulations 2020, SI 2020/1031, para 2(4).
[82] The Corporate Insolvency and Governance Act 2020, s 12.
[83] The Corporate Insolvency and Governance Act 2020 (Coronavirus) (Suspension of Liability for Wrongful Trading and Extension of the Relevant Period) Regulations 2020, SI 2020/1349, para 2(2).
[84] ibid s 2(1).
[85] Corporate Insolvency and Governance Act 2020, s 12.
[86] Insolvency Act 1986 s 213, s 246ZA.
[87] ibid s 423.
[88] Companies Act 2006, ss 171-177.
[89] ibid.

proceedings to continue trading through the suspension period, while the financial position of the company and its creditors gets worse without incurring personal liability, subject to not breaching the above conduct provisions still in force.

CIGA also introduced temporary changes to how and when creditors may present winding-up petitions and orders with retrospective effect.[90] Under normal circumstances the IA[91] authorises a company to be wound up by the court if it is unable to pay its debts.[92] CIGA suspended creditors from presenting winding-up petitions and the courts from granting winding-up orders based on unpaid statutory demands or unpaid judgment debts in the suspension period.[93] Unless the creditor or court had reasonable grounds for believing that COVID-19 had not had a financial effect on the company, or the company's inability to satisfy the statutory demand or judgment debt was not caused by COVID-19.[94] The retrospective nature of CIGA empowers the courts to reverse winding-up orders granted on, or after 27 April 2020, if the orders would not have been made had CIGA been in force and the above test applied. The suspension period which ran from 1 March 2020 to 30 September 2020[95] was subsequently extended till 31 March 2021.[96] Ultimately, in the event that COVID-19 has not had a financial effect on the company, or the company would be in financial distress regardless of

[90] Corporate Insolvency and Governance Act 2020, s 10.
[91] Insolvency Act 1986, s 122(1)(f).
[92] A company is deemed unable to pay its debts in accordance with Insolvency Act 1986, s 123(1) if, among other reasons, a creditor, to whom the company is indebted in a sum exceeding £750, serves a statutory demand requiring the company to repay the debt and the company has failed to repay that debt for three weeks thereafter. The company is cash-flow insolvent, being unable to pay its debts as they fall due, or is balance sheet insolvent, where the value of the company's assets is less than the amount of its liabilities, taking into account its contingent and prospective liabilities.
[93] Between 1 March 2020 and 30 September 2020; Corporate Insolvency and Governance Act 2020, sch 10, pt 1.
[94] Corporate Insolvency and Governance Act 2020, sch 10, paras 2(4) and 5.
[95] ibid sch 10, pt 1.
[96] First extended till 31 December 2020 by Extension of the Relevant Period Regulations 2020 (n 81) para 2. Subsequently extended by The Corporate Insolvency and Governance Act 2020 (Coronavirus) (Extension of the Relevant Period) (No 2) Regulations 2020, SI 2020/1483, para 2.

the effects of COVID-19, the company may not benefit from the temporary restriction of winding-up petitions and orders.

A number of temporary provisions introduced by CIGA were not extended or were withdrawn.[97] This reflects the UK government's aim of balancing the need to protect viable businesses which have been affected by COVID-19 with the need to protect creditor interests in relation to businesses that are not viable irrespective of COVID-19.

Permanent Changes

Part 26A Restructuring Plan

CIGA has also introduced several permanent changes to the UK insolvency regime. One such change is the introduction of a new restructuring procedure that is closely based on the pre-existing UK scheme of arrangements,[98] but which has the ability to 'cram down' dissenting classes of creditors.[99] For a company in financial distress to utilise the new Part 26A[100] restructuring plan, two conditions must be satisfied:

(1) the company must have encountered or must be likely to encounter 'financial difficulties that are affecting, or will, or may affect, its ability to carry on business as a going concern';[101] and
(2) the purpose 'must be to eliminate, reduce or prevent, or mitigate the effect of any financial difficulties'.[102]

Mokal notes that 'financial difficulties', a term forming part of both conditions, is not defined, but presumably takes its meaning from the first condition, 'affecting … the company's ability to carry on

[97] The Corporate Insolvency and Governance Act 2020 (Coronavirus) (Early Termination of Certain Temporary Provisions) Regulations 2020 SI 2020/1033.
[98] Companies Act 2006, pt 26.
[99] ibid s 901G.
[100] ibid pt 26A, inserted by Corporate Insolvency and Governance Act 2020, sch 9 para 1.
[101] ibid pt 26A, ss 901A(2).
[102] ibid pt 26A, ss 901A(3).

business as a going concern', though this is unclear.[103] Furthermore, there is no necessary connection between a company's inability to meet its liabilities as they fall due and whether its assets are currently deployed in their optimal use.[104] Ultimately, preservation of a company as a going concern may not maximise the value of its assets, depending on whether the company is economically distressed. If economically distressed, preservation of a company as a going concern may exacerbate rather than mitigate its financial difficulties.[105] The new restructuring plan seemingly represents a shift in the law towards keeping a business together, rather than asset stripping. A restructuring plan must focus on the business as a going concern,[106] aiming to enable a company to trade out of its difficulties without harassment and dismemberment by creditors.[107] A restructuring plan does not need to seek to restore the financial ability of the company; only to mitigate the effect of those difficulties upon an unspecified someone, be it some or all creditors or even the company itself. In this regard a Part 26A restructuring plan is its 'very own Duke of York, the two conditions look set to march ten thousand companies to the top of the going concern hill, only to march them down again' when they succumb to their economic distress.[108]

Historically one of the limitations of a UK Part 26 Scheme[109] is the inability for dissenting classes to be bound by the vote of other classes. Part 26A has addressed this gap in the UK's restructuring tool kit by enabling 'cross-class cram down' of dissenting classes.[110] This comes at the expense of having financial entry requirements,[111] unlike the original government proposal.[112] Like a scheme of arrangements, Part 26A is a court process, whereby an initial court hearing is held to

[103] Riz Mokal, 'The difficulties with "financial difficulties": the threshold conditions for the new Pt 26A process' (2020) 35(10) BJIB&FL 662.
[104] ibid 663.
[105] ibid.
[106] ibid 662.
[107] Riz Mokal, 'The two conditions for the Pt 26A cram down' (2020) 35(11) BJIB&FL 730.
[108] Mokal, 'The difficulties with "financial difficulties"' (n 103) 664.
[109] Companies Act 2006, pt 26.
[110] ibid pt 26A, ss 901F and G.
[111] ibid s 901A(2). For discussion see Mokal, 'The difficulties with "financial difficulties"' (n 103) 664.
[112] BEIS, *Insolvency and Corporate Governance* (n 79).

determine class composition and convene meetings of the creditors and shareholders which will be subject to the restructuring plan. Following the relevant class meetings, there is a second court hearing where the court has the discretion to sanction the restructuring plan after considering, amongst other matters, whether all procedural formalities were complied with and whether the plan is 'fair'.[113] If the plan has been agreed by at least 75% in value of class creditors the court can sanction the restructuring plan under the CA section 901C. Where less than 75% in value of class creditors approve the plan, it is open to the court to sanction the restructuring plan under CA section 901G, thereby 'cramming down' dissenting classes.

Dissenting classes can only be subject to 'cram down' under CA section 901G where two conditions are met. Condition A[114] is met where the court is satisfied that members of the dissenting class would not be any worse off than they would be in the event of the 'relevant alternative'.[115] The concept of 'relevant alternative' is perhaps the key safeguard against abuse of the 'cram down' mechanism.[116] The principle of 'relevant alternative' has a venerable lineage, going back to the Court of Appeal's decision in *Re English, Scottish, and Australian Chartered Bank*.[117] It has been endorsed by the Cork Report,[118] in the context of class formation by Lewison J in *Re British Aviation Insurance Co Ltd*,[119] and for the "vertical comparator"[120] in the context of an unfair prejudice challenge to a CVA by Norris J in *Discovery Ltd v Debenhams Ltd*.[121] Condition B requires the restructuring plan to have been approved by at least one class whose members would in the

[113] Mokal, 'The two conditions for the Pt 26A cram down' (n 107).
[114] Companies Act 2006, pt 26A, s 901G(3).
[115] ibid ss 901G(3)- and (4). Condition A requires the court to be satisfied that no member of the dissenting class would be worse off under the plan than they would be in the event of the 'relevant alternative', which is whatever the court considers most likely to occur in relation to the company if the plan were not sanctioned.
[116] Mokal, 'The two conditions for the Pt 26A cram down' (n 107) 730.
[117] *Re English, Scottish, and Australian Chartered* Bank [1893] 3 ch 385, 413 (Lindley LJ) and 414-5 (Lopes LJ).
[118] explained by Etherton J in; *Prudential Assurance Co Ltd v PRG Powerhouse Ltd* [2007] EWHC 1002 (Ch) [78]-[79].
[119] *Re British Aviation Insurance Co Ltd* [2005] EWHC 1621 (Ch) [88].
[120] Comparison between the projected outcome of the CVA with the projected outcome of a realistically available alternative process, usually liquidation.
[121] *Discovery Ltd v Debenhams Ltd* [2019] EWHC 2441 (Ch) [12].

relevant alternative receive a payment, or have a genuine economic interest in the company.[122] This condition appears inspired by Chapter 11 of the US Bankruptcy Code.[123]

A key weakness identified by Mokal[124] is the threshold conditions in CA section 901A which determine whether a company can make use of the Part 26A restructuring plan by proposing a 'compromise or arrangement.'[125] The definition of 'compromise' has in the past been considered judicially. It has been held that the power to compromise a right arises only when there is a dispute or difficulty concerning a right 'which cannot be got over without some arrangement'.[126] The term implies 'some element of accommodation on each side' and not surrender.[127] While an arrangement does not require the existence of a dispute or difficulty,[128] it does imply some element of give and take.[129] However, the jurisprudence on 'horizontal comparators'[130] is potentially intrusive for the purposes of an unfair prejudicial challenge to a restructuring plan on the basis that differential treatment amongst claimants is not justifiable.[131] If successfully challenged on this ground, the court would appear to be deprived of the ability to 'cram down' the restructuring plan.[132]

The first case to engage with the new Part 26A restructuring plan was *Re Virgin Atlantic Airlines Ltd*, a simple case that did not implicate any of the issues raised above. For this reason, Trower J was

---

[122] Companies Act 2006, s 901G(5).
[123] 11 US Code § 1129(a)(10) impaired class acceptance requirement. For discussion see Mokal, 'The two conditions for the Pt 26A cram down' (n 107) 731.
[124] Mokal, 'The difficulties with "financial difficulties"' (n 103) 622.
[125] Companies Act 2006, s 901A(3).
[126] *Sneath v Valley Gold, Ltd* [1893] 1 ch 477 (CA), 493-494 (Lindley LJ).
[127] *Re National Farmers' Union Development Trust* [1972] 1 WLR 1548 (ChD) 1555C-D (Brightman J).
[128] In *Re Guardian Assurance Company* [1917] 1 ch 431 (CA) 449-450 (Warrington LJ).
[129] ibid 450. See also *Re National Farmers* (n 127).
[130] The comparison of treatment of creditors in an arrangement relative to one and other.
[131] *Prudential Assurance Co Ltd v PRG Powerhouse Ltd* [2007] EWHC 1002 (Ch), [75] and [86]-[91] (Etherton J); and
Discovery Ltd v Debenhams Ltd [2019] EWHC 2441 (Ch), [12] and [102]-[110] (Norris J).
[132] Mokal, 'The two conditions for the Pt 26A cram down' (n 107) 732.

able to find in the initial hearing that the two conditions were met.[133] Snowden J sanctioned the restructuring plan a month later with no dissenting creditors.[134] The subsequent landmark case of *Re Deepocean 1 UK* Ltd[135] was not so simple. Trower J approved the first 'cram down' restructuring plan under CA section 901F where CA section 901G was engaged because only 64.6% of Deepocean Subsea Cables Ltd's 'Other Plan Creditors' in value approved the restructuring plan.[136] Marking a historic change to the UK insolvency regime and paving the way for future 'cram down' restructuring plans.

Ipso Facto *Clauses*

CIGA broadened the existing ban on *ipso facto* clauses contained within the IA[137] which would otherwise permit suppliers of goods and services to terminate or vary supply to companies due to the commencement of insolvency and restructuring procedures.[138] The extension of the *ipso facto* ban had first been proposed in the Insolvency Services 2016 'Review of Corporate Insolvency Framework'.[139] CIGA prohibits such clauses, subject to certain exceptions, in order to assist distressed companies as a going concern. This represents a shift from freedom of contract to prioritising corporate rescue.

Prior to CIGA, the IA[140] aimed to assist with the rescue of an insolvent company as a going concern by protecting a business's operational capabilities in situations of financial distress through preserving the supply of utilities. The effect of the modification to IA

[133] *Re Virgin Atlantic Airway Ltd* [2020] EWHC 2191 (Ch) [37]-[39].
[134] *Re Virgin Atlantic Airway Ltd* [2020] EWHC 2376 (Ch) [41]-[43].
[135] *Re Deepocean* 1 UK Ltd (n 7).
[136] ibid [5].
[137] Insolvency Act 1986, s 233A.
[138] A company in financial distress does not necessarily have to be incorporated in the UK; the company may be utilising the relevant insolvency procedure due to its COMI (Centre of main interest) being based in the UK or if there is a sufficient connection to the UK (with the test varying depending on which insolvency proceeding is being utilised).
[139] Insolvency Service, *A review of the Corporate Insolvency Framework* (n 41) 60-61.
[140] Insolvency Act 1986, s 233.

section 233A[141] and introduction of section 233B[142] is that *ipso facto* clauses now cease to affect contracts for the supply of 'essential' goods and services. Hotton and Norris note that a supplier can terminate the contract, through the company or relevant officer-holders consent, or with permission from the court, if the continuation of supply would cause 'hardship' on the supplier.[143]

Hotton and Norris have also identified several issues non-financial suppliers may face as a result of the new *ipso facto* clause rules.[144] Non-financial suppliers of goods and services, mostly trade suppliers, will now be unable to terminate or 'do any other thing'[145] on the commencement of an insolvency procedure. However, they will retain the right to terminate supply concerning any non-insolvency related events contained in the contract, where this occurs after the commencement of the insolvency procedure.[146] The carve-outs to IA section 233B are provided for in Schedule 4ZZA and include certain financial contracts and persons involved in financial services.[147] Notably, to a large extent, the carve-outs mirror the regime under the US Bankruptcy Code. Crucially, the financial contracts exclusion lacks clarity as to whether it covers all loan transaction documentation.[148] Due to the carve-out for financial contracts, the distressed company will not be able to prevent financial creditors from taking away committed funds including working capital facilities, enabling creditors to revoke capital at a time where it is vital for the survival of the company. At the time of writing, there was a

[141] ibid s 233A amended by Corporate Insolvency and Governance Act 2020.
[142] ibid s 233B inserted by Corporate Insolvency and Governance Act 2020.
[143] ibid s 233B(5).
[144] Lorna Hotton and Jasmine Norris 'UK Corporate Insolvency and Governance Act: effects on ipso facto clauses' 2020 35(8), BJIB&FL 550, 511
[145] Insolvency Act 1986, s 233A(8)(a).
[146] ibid ss 233A(2)(b) and 233B(3)(b).
[147] In summary loan agreements, hedging arrangements and other types of financial contract are carved-out of the application of Insolvency Act 1986, s 233B and certain entities (eg investment banks and insurance companies) are excluded from the effects of Insolvency Act 1986, s 233B regardless of whether they are the insolvent entity or the supplier.
[148] Hotton and Norris (n 144), which notes that on a strict reading of the new Insolvency Act 1986, sch 4ZZA, intercreditor agreements would not be excluded from the prohibition with it unclear whether intercreditor agreements are agreements for the supply of services.

temporary exclusion on *ipso facto* clauses for small suppliers in the UK subject to CIGA section 15(1) which had been extended until 30 March 2021.[149]

Overall, IA section 233B[150] will be a welcome addition to the insolvency toolkit for distressed companies and closely aligns the UK insolvency regime with that of the US. The measures prioritise distressed debtor companies and financial parties over the interests of general suppliers. Implementation of a system more akin to that included in the US Bankruptcy Code[151] would have facilitated a fairer system. Additionally, where liquidation is the only outcome, suspension of *ipso facto* clauses is not justifiable, which makes the inclusion of liquidation on the list of 'relevant insolvency procedures'[152] problematic.[153]

New Moratorium Procedure

A period of moratorium has existed as an integral aspect of mechanisms such as UK administration and US Chapter 11 bankruptcy for some time. Moratoria are traditionally regarded as having two benefits: mitigating the 'common pool' problem;[154] and the 'anti-commons' problem, in that it can block actions by individual creditors who are seeking to frustrate the wishes of the majority.[155] However, these benefits occur at the expense of party autonomy, since minority creditors are prevented from exercising their contractual rights.

[149] Extension of the Relevant Period Regulations 2020 (n 81) para 2(2)(a).
[150] Insolvency Act 1968, s 233B.
[151] US Bankruptcy Code § 365.
[152] Insolvency Act 1986, s 233B(1).
[153] Jennifer Payne, 'An Assessment of the UK Restructuring Moratorium' (2021) University of Oxford Faculty of Law <https://ssrn.com/abstract=3759730> accessed 4 April 2021.
[154] Thomas H Jackson, *The Logic and Limits of Bankruptcy Law* (Beard Books 2001)
[155] Douglas Baird and Robert Rasmussen, 'Anti-bankruptcy' (2010) 119 Yale L J 648, Rolef de Weijs, 'Harmonisation of European insolvency law and the need to tackle two common problems: common pool and anticommons' (2012) 21 International Insolvency Review 67.

Until 2020 only two options for a moratorium existed in the UK for those involved in debt restructuring: the automatic moratorium attached to administration;[156] and that available for CVAs related to small companies.[157] However, neither moratorium was particularly beneficial from a debt restructuring perspective. Administration lacks many of the required components for a successful debt restructuring mechanism.[158] In particular, it is not a debtor-in-possession procedure,[159] and as an insolvency mechanism it brings with it the stigma of insolvency operating at a relatively late stage in the process compared to other available mechanisms, such as schemes of arrangement.[160] CVAs are perhaps more valuable as a debt restructuring mechanism[161] as they are a debtor-in-possession procedure, can operate at an earlier stage than administration, and can 'cram down' the rights of the minority of unsecured creditors. However, they are limited by the fact secured and preferential creditors cannot be bound without their consent.[162]

CIGA inserted Part A1 into the IA.[163] In doing so, it provided a standalone restructuring moratorium that is not a precursor to an insolvency process, although it can be used in that way. To utilise the new moratorium a company must not have been subject to a moratorium or certain formal insolvency procedures in the last 12 months,[164] although these provisions were suspended until 30 March

[156] Insolvency Act 1986, sch B1, paras 42-3.
[157] ibid s 1A and sch A1.
[158] Jennifer Payne, 'Debt restructuring in English Law: Lessons from the US and the need for reform' (2014) Law Quarterly Review 282.
[159] For discussion of debtor-in-possession versus creditor-in-possession procedures see David Hahn, 'Concentrated Ownership and Control of Corporate Reorganizations' (2004) JCLS 117, Sefa Franken, 'Creditor- and Debtor-Oriented Corporate Bankruptcy Regimes Revisited' (2004) 5 EBOR 645.
[160] The entry into administration will generally trigger cross-default clauses or contractual provisions for acceleration of liability, termination of the contract or the cessation of intellectual property rights by contrast the entry into other debt restructuring mechanisms, such as schemes of arrangement, does not tend to operate as such a trigger.
[161] For example, *Discovery Ltd v Debenhams Retail Ltd* [2019] EWHC 2441 (Ch).
[162] Insolvency Act 1986, s 4(3).
[163] Corporate Insolvency and Governance Act 2020, ss 1-6 and schs 1-8, which introduce new provisions into the Insolvency Act 1986 (for Great Britain) and into the Insolvency (Northern Ireland) Order 1989, SI 1989/2405 (NI 19).
[164] Extension of the Relevant Period Regulations 2020 (n 81) para 2.

2021.[165] In contrast to the moratorium attached to administration, the new moratorium is a debtor-in-possession device allowing directors to continue to run a company subject to the appointment of a licensed insolvency practitioner (monitor) and certain other restrictions.[166] The objective of the moratorium is to provide protection to the company from adverse creditor action for a short period whilst the director's attempt to restructure and rescue the business. However, CIGA leaves a number of key issues undefined and relies heavily on court interpretation to develop the functionality of the new moratorium procedure. The test inserted by CIGA requires a monitor to state that in their view it is likely that a moratorium would result in the rescue of the company as a going concern.[167] The court has been left to establish the boundaries for this test, with the danger that this could set a high bar, resulting in the exclusion of numerous financially distressed companies where there is a chance of rescue and the risk of abuse is low.

The need for creditors to be constrained from action is well understood in the context of insolvency law as a means of keeping a business and its assets together so that they can be sold for the highest possible price.[168] The US Chapter 11 bankruptcy regime is often regarded as the 'gold standard' in terms of a mechanism for managing corporate reorganisation.[169] Immediately upon the filing of a bankruptcy petition under the US system a moratorium automatically arises and stays all litigation and prevents the enforcement of judgments and of security without leave of the court.[170] Debtors are given the exclusive right to formulate a plan of reorganisation for 120 days from the date of filing. This period can be extended if sufficient

[165] ibid para 2(b).
[166] Insolvency Act 1986, pt A1.
[167] ibid, pt A1, ch 3, s A6(1)(e).
[168] Thomas H Jackson, 'Bankruptcy, Non-bankruptcy Entitlements, and the Creditors' bargain" (1982) 91 Yale Law Journal 857; and
Thomas H Jackson and Robert E Scott, 'On the Nature of bankruptcy: An Essay on Bankruptcy Sharing and the Creditors' Bargain' (1989) 75 Virginia Law Review 155.
[169] Elizabeth Warren and Jay Westbrook, 'The Success of Chapter 11: A Challenge to the Critics' (2009) 107 Michigan Law Review 603.
[170] 11 US Code § 362(a). The stay is effective during the entire time the case is pending but creditors and other parties may make motions to lift or modify the stay, § 362 (d).

reasons are established, up to a maximum of 18 months from the filing date.[171] The US regime restricts the ability of creditors to terminate contracts on the basis of insolvency[172] and allows the debtor to 'cherry pick' among outstanding contracts, subject to certain safeguards. Despite the Government's aspirations in this regard, the new moratorium seems unlikely to benefit small and medium sized enterprises (SMEs). A great deal of work has been done in recent years to move away from a one size fits all approach and to consider solutions to deal with financially distressed SMEs.[173] More surprising perhaps is the fact that the moratorium is unlikely to be available, or valuable, for many larger companies seeking to restructure their debt. This is due to the very short time period for the moratorium (only 40 days without court or creditor approval),[174] which is a short period in which to conclude a complex restructuring.[175] In particular, the wide carve-out for any company subject to a capital market arrangement will exclude numerous businesses which have bond financing.[176] There is no doubt that restructuring mechanisms, including moratoria, can be misused by companies and powerful financial creditors.[177] However, while the concerns about protecting the minority from abuse are important, the balance may have tipped too far in these provisions, rendering the restructuring moratorium less valuable for companies than had been hoped and anticipated.[178] Additionally, the utility of the new moratorium is limited by the number of pre-moratorium debts that are required to be paid,[179] including the monitor's remuneration and expenses, goods or services supplied during the moratorium, rent (for the moratorium period), wages, salary, redundancy payments and amounts payable in respect of debts or liabilities arising under a contract or other instrument

[171] 11 US Code § 1121.
[172] 11 US Code § 365(e) and § 541(c).
[173] UNCITRAL Working Group V (Insolvency Law) 'Insolvency of Micro, Small and Medium-Sized Enterprises' (2018) UN Doc A/CN.9/WG.V/WP.163;
Riz Mokal, Ronald Davis, Alberto Mazzoni, and others, *Micro, Small, and Medium Enterprise Insolvency: A Modular Approach* (OUP 2018).
[174] Insolvency Act 1986, pt 1A, ch 3.
[175] Payne, 'An Assessment of the UK Restructuring Moratorium' (n 153) 19.
[176] ibid and Insolvency Act 1986, sch 4ZZA.
[177] For a discussion of this in the context of US Chapter 11 see Douglas Baird and Robert Rasmussen, 'Chapter 11 at Twilight' (2003) 56 Stanford Law Review 673.
[178] Payne, 'An Assessment of the UK Restructuring Moratorium' (n 153) 9.
[179] ibid.

involving financial services.[180] Overall, the necessary balance has not been achieved and the constraints and limitations placed on the moratorium to protect creditors will significantly restrict its utility.[181]

Return to Crown Preference

After two decades Crown preference has been partially reintroduced by the Finance Act 2020 section 98, inserting a new category of secondary preferential debts into the IA.[182] Despite widespread criticism from the restructuring and insolvency industry,[183] the government implemented the change without incorporating amendments that would have tempered its impact. On 29 October 2018, the Chancellor of the Exchequer announced in the budget, without any prior consultation, the re-introduction of Crown preference from 6 April 2020 for certain taxes collected by businesses 'on behalf of' another.[184] The Chancellor justified the reintroduction stating that those monies are 'held in trust by the business … to fund public services' which is legally incorrect.[185] As secondary preferential debts, the newly preferential tax debts will rank behind fixed charge holders' claims and ordinary preferential debts but ahead of floating charge holders and unsecured creditors.[186] HMRC remains an unsecured creditor for other unpaid taxes such as corporation tax and employer national insurance contributions.

This may lead to the new Crown preference significantly eroding and, in some cases, wiping out floating charge realisations as well as any returns to unsecured creditors. This is particularly true in the current climate where considerable VAT deferrals have been granted by HMRC as part of COVID-19 support measures. Government figures estimate that £33.5 billion of VAT due between

[180] Insolvency Act 1986, pt A1, ch 4, s A18(3).
[181] Payne, 'An Assessment of the UK Restructuring Moratorium' (n 153).
[182] Insolvency Act 1986, sch 6 para 15D.
[183] *Protecting your taxes in insolvency: Summary of responses* (HMRC 2019).
[184] *Budget 2018: Protecting your taxes in insolvency* (HM Treasury 2018).
[185] Inga West & Rebecca James, 'The return of Crown preference: a backward step' 2021 34(1) Insolv Int 2021 3.
[186] Insolvency Act 1986, s 175.

20 March 2020 and 30 June 2020 has been deferred by HMRC.[187] At the time of writing, debtors will have until March 2022 to repay the deferred VAT.[188] The re-introduction of Crown preference has effectively transferred the insolvency credit risk in the VAT deferrals from HMRC to floating charge holders and unsecured creditors.

The re-introduction of Crown preference may also limit the usefulness of CVAs, as preferential debts cannot be compromised in a CVA without the preferential creditor's consent. Historically, HMRC has not been supportive of CVAs and as a result, we may be witnessing their demise.[189]

Part Conclusion

The new measures introduced by CIGA are welcome tools to the UK restructuring and insolvency arsenal. *Pro tempore* financially distressed companies will benefit from measures designed to promote their survival, however, in the future functionality and utility of the permanent measures may be somewhat squandered by the preference to protect corporate creditors and the re-introduction of Crown preference.

[187] 'HMRC coronavirus (COVID-19) statistics' (*HMRC*) available at <www.gov.uk/government/collections/hmrc-coronavirus-covid-19-statistics#vat-paymentsdeferral-scheme> accessed 5 April 2021.
[188] 'VAT deferred due to coronavirus (COVID-19)' (*HMRC*, 26 March 2020) <www.gov.uk/guidance/deferral-of-vat-payments-due-to-coronavirus-covid-19> accessed 13 January 2021.
[189] Peter Walton, Chris Umfreville and Lezelle Jacobs, 'A Snapshot of company voluntary arrangements: Success, failure and proposals for reform' 2020 29(2) Int Insolv Rev 267.

A comparative analysis with Germany and the USA

Introduction

In the wake of COVID-19, economists estimate UK GDP fell by 9.9%,[190] the largest annual fall in 300 years,[191] with global GDP falling by 4.3%.[192] States have adopted measures to alleviate the financial pressure companies find themselves under as a result of the economic contraction. The corporate insolvency[193] responses of Germany and the US have differed greatly. Germany has responded to COVID-19 with a number of legislative changes to its insolvency regime, including the introduction of a new restructuring process. The US has implemented a single legislative change by increasing the threshold for businesses to access streamlined bankruptcy protection to $7,500,000,[194] focusing instead on expansionary fiscal and monetary policy.[195] This part shall analyse the measures and approach adopted by Germany and the US, and analyse the UK-US cross-border case of *Re Virgin Atlantic Airways Ltd.*[196] In doing so the approach adopted by the UK shall be compared to that of Germany and the US.

Virgin Atlantic Restructuring

Airline failures in Europe are not uncommon;[197] the fragmented airline market suffers from oversaturation and a reliance on

---

[190] *GDP first quarterly estimate, UK: October to December 2020* (Office for National Statistics 2021).

[191] 'A millennium of macroeconomic data' (*Bank of England*) <www.bankofengland.co.uk/statistics/research-datasets> accessed 25 March 2021.

[192] *January 2021: Global Economic Prospects* (World Bank 2021) 3-4.

[193] In Germany and the US there is no distinction in name between personal bankruptcy and corporate insolvency as there is in the UK.

[194] Coronavirus Aid, Relief, and Economic Security Act, modifying Small Business Reorganization Act of 2019.

[195] Contained within The Coronavirus Aid, Relief, and Economic Security Act of 2020; The American Rescue Plan Act of 2021.

[196] *Re Virgin Atlantic Airways Ltd* (n 6).

[197] Victoria Tozer-Pennington, The Aviation Industry Leaders Report 2021: Route to Recovery (Aviation News Limited 2021) 19, and 'Europe's airline industry is consolidating' *The Economist* (London, 27 April 2019) <www.economist.com/business/2019/04/27/europes-airline-industry-is-consolidating> accessed 2 April 2021.

international travel.[198] This meant that national lockdowns and international travel restrictions imposed due to COVID-19 severely impacted European airlines.[199] Flybe was an early victim of the highly competitive European air travel market,[200] ceasing to operate in March 2020.[201] Flybe had been in financial difficulties in the years prior to the pandemic and was subsequently sold to Connect Airways Ltd, a consortium of Virgin Atlantic Airways (VAA), Stobart Aviation and Cyrus Capital Partners whom intended to rebrand Flybe as 'Virgin Connect'. Connect Airways offered Flybe a £20 million bridging facility to keep trading, conditional on a modification of terms from the credit card companies. That could not be done, so Flybe sold Connect its trading subsidiaries under a sale and purchase agreement. The scheme offered 1p per share to shareholders and was sanctioned by the court.[202] Ultimately when VAA was not willing to make more money available due to the absence of any prospect of recovery in light of COVID-19 Flybe was wound up on 24 February 2021.[203] This marked the beginning of VAA's financial difficulties.

The UK government has supported airlines through the furlough payment scheme, with larger companies eligible for the Covid Corporate Financing Facility (CCFF) scheme, designed to boost liquidity and bridge COVID-19 disruption to cash flows through the purchase of short-term debt in the form of commercial paper.[204] British Airways, easyJet, Jet2, Ryanair and Wizz Air, have all raised funds in the UK via CCFFs. However, VAA does not qualify for the CCFF scheme. An important source of government support for the

[198] The Economist, 'Europe's airline industry is consolidating' (n 197).
[199] Aviation News, 'The Aviation Industry Leaders Report 2021' (n 197) 14.
[200] 'Flybe Group Limited Petitions to Wind Up' *The London Gazette* (London, 15 February 2021) 2628.
[201] Letter from S J Edel, Joint Administrator, to all known unsecured creditors of Flybe Limited (in Administration) (5 August 2020).
[202] *Re Flybe Group Plc* [2019] EWHC 631 (Ch).
[203] The Gazette, 'Flybe Group Limited Petitions to Wind' (n 200).
[204] 'Covid Corporate Financing Facility (CCFF): information for those seeking to participate in the scheme' (*Bank of England*, 20 March 2020) <www.bankofengland.co.uk/news/2020/march/the-covid-corporate-financing-facility%20> accessed 14 May 2022.

aviation sector in the UK is the Export Development Guarantee,[205] which helps large UK exporters access high-value loan facilities for general working capital or capital expenditure purposes.[206] In June 2020, Rolls-Royce was the first company to access this funding and secured a £2 billion five-year term loan under this facility, which was 80%-backed by UK Export Finance. This deal has since been replicated for British Airways and easyJet. However, VAA was again ineligible.[207]

On 14 July 2020 VAA announced a five-year restructuring plan, the core of which was a refinancing package worth c.£1.2 billion, in addition to cost savings of c.£280 million and c.£880 million rephasing and financing of aircraft deliveries.[208] On 4 August 2020, Trower J approved VAA's application for an order convening meetings of certain of its creditors for the purpose of approving a restructuring plan under CA section 901A. VAA Ltd is 51% owned by Virgin Investments Ltd, a UK company and 49% owned by Delta Air Lined Inc, a major US airline, in addition to having US subsidiaries.[209] This meant VAA could also access the US insolvency regime; in particular, Chapter 15 bankruptcy protection due to having assets in the US.[210] VAA sought Chapter 15 recognition of the restructuring plan due to the relief which would be granted over the company's assets, protecting VAA from its creditors.[211] On 4 August 2020, the UK Part 26A restructuring plan was recognised under Chapter 15 and relief granted over VAA's assets.

---

[205] 'Export Development Guarantee' (*UK Export Finance*, 21 July 2020) <www.gov.uk/guidance/export-development-guarantee> accessed 22 March 2021.
[206] *Business Plan 2020-2024* (UK Export Finance 2020).
[207] Aviation News, 'The Aviation Industry Leaders Report 2021' (n 197) 14.
[208] 'A Solvent Recapitalisation of Virgin Atlantic' (*Virgin Atlantic Airways*, 14 July 2020) <https://corporate.virginatlantic.com/gb/en/media/press-releases/a-solvent-recapitalisation-of-virgin-atlantic.html> accessed 21 March 2021.
[209] Aviation News, 'The Aviation Industry Leaders Report 2021' (n 197) 46.
[210] 11 US Code: Chapter 15 Ancillary and Other Cross-Border Cases § 1502,11 US Code § 1517.
[211] *Re Virgin Atlantic Airways Ltd* (n 133) [66].

The United States

At the time of writing, the US had implemented limited change to its bankruptcy law in response to COVID-19. This might be because the US bankruptcy regime is often taken to be the gold standard by practitioners, with Chapter 11 the inspiration for debt restructuring mechanisms across the globe.[212] The approach of the US has somewhat relied on the strength of its present regime, Chapter 7 liquidation bankruptcy, Chapter 11 reorganizational bankruptcy, and Chapter 15 cross-border bankruptcy. On 27 March 2020, The Coronavirus Aid, Relief, and Economic Security Act of 2020 (CARES) was signed into law by President Donald Trump and with it, $2 trillion in stimulus funding. On 21 December 2020 a further $900 billion end-of-the-year COVID-19 stimulus bill was passed, and most recently The American Rescue Plan Act of 2021, a $1.9 trillion stimulus package signed into law by President Joe Biden. The total allocated at the time of writing by the US to alleviate the impact of COVID-19 exceeds $7.41 trillion.[213]

Part of the US approach to mitigate the financial distress faced by companies has been a $500 billion fund controlled by the Federal Reserve made available through a government lending programme. Additionally, specifically to target SMEs, a $670 billion fund was made available through the Small Business Administration Loan Program and the Paycheck Protection Program. Under CARES, the threshold allowing businesses to take advantage of the streamlined bankruptcy protections available under the Small Business Reorganization Act of 2019 was raised from approximately $2,725,625 to $7,500,000.[214] This represents the sole legislative change to the US

---

[212] For other examples see, Spanish Law 38/2011 of October 2011 and Law 4/2014 of 7 March, discussed in Ignacio Tirado, 'Scheming against schemes: A New framework to deal with Business Financial Distress in Spain' [2018] European Company and Financial Law Review 516; and
Wet homologatie onderhands akkoord (Act on the Confirmation of Private Plans) (Netherlands).

[213] 'The Federal Response to COVID-19' (*US Government*) <https://datalab.usaspending.gov/federal-covid-funding/> accessed 2 April 2021.

[214] The Coronavirus Aid, Relief, and Economic Security Act of 2020 modifying Small Business Reorganization Act of 2019, subchapter V.

bankruptcy regime made in response to COVID-19. The change applies for one year from the effective date of CARES and though this may not substantially modify the US Bankruptcy Code, the temporary change may result in an increase of filings under the streamline process where companies would otherwise have filed for Chapter 11-proceedings.

In contrast to the UK, the US has no requirement that the debtor is insolvent or near insolvency in order to apply for Chapter 11 protection. The process is an instrument for debtor relief, not a remedy for creditors.[215] This is perhaps why there has been no modification of the present US bankruptcy regime: Chapter 11 already provides up to 18 months of moratorium protection for companies.[216]

The US has enacted limited COVID-19 specific insolvency legislation, only modifying the threshold for proceedings under the Small Business Reorganization Act of 2019. However, this may be because the regime in place is considered robust, with any weakness alleviated by the expansionary fiscal and monetary policy adopted and the ability of the regime to recognise foreign procedures.[217]

Germany

Bankruptcy proceedings in Germany are governed by a single codified legal document, Insolvenzordnung (InsO).[218] There is a single point of entry from which financially distressed companies are assessed, and the most appropriate course of action decided.[219] The main purpose of German insolvency proceedings is to provide a method for the realisation of a debtor's assets in a manner that treats

[215] Paul B Lewis 'Corporate Rescue Law in the United States' in Katarzyna Gromek Broc; and
Rebecca Parry, C*orporate Rescue: An Overview of Recent Development* (Kluwer Law International 2006) 333.
[216] 11 US Code § 1121.
[217] For example, the US Chapter 15 protection was granted over the novel pt 26A restructuring plan introduced by the UK during COVID-19
[218] Insolvenzordnung (Insolvency Code).
[219] Finch and Milman (n 4) 239.

all unsecured creditors equally.[220] On 27 March 2020 COVID-19-Insolvenzaussetzungsgesetz (COVInsAG)[221] was introduced into German federal law. COVInsAG introduced a number of measures specifically intended to help companies in financial difficulty and to reduce the number of companies forced to enter into insolvency proceedings. In addition to implementing measures to combat the impact of COVID-19 Germany has also introduced a new restructuring process. On 19 September 2020, the German Ministry of Justice published a draft law[222] transposing the EU Restructuring and Insolvency Directive 2019[223] in addition to other new provisions.

COVID-19-Insolvenzaussetzungsgesetz

Adopted on 27 March 2020, with retrospective effect from 1 March 2020, COVInsAG § 1 suspends the obligation of a director to file for insolvency in accordance with InsO § 5[224] and the German Civil Code[225] until 30 September 2020.[226] As of 1 October 2020, the obligation to file for bankruptcy in the case of illiquidity (*Zahlungsunfähigkeit*) cash flow insolvency was reinstated, whereas the obligation to file for insolvency in case of COVID-related over-indebtedness (*Überschuldung*) was extended until 31 December 2020. Similar to the UK, suspension does not apply to companies that have become insolvent not as a result of the COVID-19 pandemic, or where

---

[220] Michael Schillig, 'The Transition from Corporate Governance to Bankruptcy Governance Convergence of German and US Law' (2010) 7 ECFR 116.

[221] *Gesetz zur vorübergehenden Aussetzung der Insolvenzantragspflicht und zur Begrenzung der Organhaftung bei einer durch die COVID-19-Pandemie bedingten Insolvenz* (COVInsAG) (Law on the temporary suspension of the obligation to file for insolvency and the limitation of directors' liability in the event of insolvency caused by the COVID-19 pandemic).

[222] Entwurf eines Gesetzes zur Fortentwicklung des Sanierungs- und Insolvenzrechts (Draft of a law for the further development of the restructuring and insolvency (Sanierungsrechtsfortentwicklungsgesetz – SanInsFoG).

[223] Directive 2019/1023 of the European Parliament and of the Council of 20 June 2019 on preventive restructuring frameworks, on discharge of debt and disqualifications, and on measures to increase the efficiency of procedures concerning restructuring, insolvency and discharge of debt, and amending Directive (EU) 2017/1132 [2019] OJ L 172/18.

[224] *Insolvenzordnung* (Insolvency Code).

[225] *Bürgerliches Gesetzbuch* (Civil Code) § 42 *Insolvenz* (Insolvency).

[226] COVInsAG (n 221) § 1(2) *Aussetzung der Insolvenzantragspflicht* (Suspension of the obligation to file for insolvency).

there is no prospect of a company overcoming existing cash flow insolvency.[227] There is a rebuttable presumption that if the company was not insolvent on 31 December 2019, the grounds for insolvency have arisen from the effects of COVID-19.[228] The burden of proof on whether these two exceptions apply lies with the party claiming that there has been a breach of the obligation to file for insolvency.

On 17 December 2020 and 26 January 2021, the German parliament enacted laws that further suspended the obligation to file for insolvency from 1 January 2021 until 30 April 2021[229] for companies which have applied for German governmental support under the COVID-19 relief programs between 1 November 2020 and 28 February 2021 or are eligible for such governmental support but have not yet filed for it, provided that such application is successful and sufficient to cure the insolvency.[230]

Under German law, directors can become personally liable for payments made after the company becomes insolvent.[231] In order to support directors of companies in financial distress COVInsAG § 2 suspended liability for payments made after insolvency in the ordinary course of business where such payments are made to maintain or resume business operations or to implement restructuring measures. COVInsAG also contains certain privileges for loans that are made during the suspension period, even if the respective company is not insolvent.[232] Newly granted loans will be protected in order to motivate lenders to provide additional liquidity to companies. Repayments of loans granted during the applicable suspension period, as well as the secured collateral for such loans, cannot be challenged in a subsequent insolvency.[233] This only applies to new loans which will not be considered disadvantageous to creditors, the mere extensions of pre-existing loans will not be

[227] ibid § 1 Suspension of the obligation to file for insolvency.
[228] ibid.
[229] SanInsFoG (n 222).
[230] COVInsAG (n 221) § 1.
[231] Insolvenzordnung § 1 Ziele des Insolvenzverfahrens (Objectives of the insolvency proceedings).
[232]COVInsAG (n 221) - § 6 *Erleichterter Zugang zum Schutzschirmverfahren* (Easier access to protective shield proceedings).
[233] SanInsFoG (n 222) § 12.

protected. Such loans will also not be subject to subordination in insolvency proceedings pursuant to InsO § 39(1).

COVInsAG also provides protection for transactions[234] which assist or enable another party to take security or obtain satisfaction of an obligation during the suspension period from being challenged in subsequent insolvency proceedings.[235] However, this privilege does not apply if the third party was aware that the debtor's restructuring and financing efforts were not suitable to cure an existing cash flow insolvency. Under German law a lender may become liable if a restructuring loan is insufficient to achieve a turnaround, and instead merely delays the filing for insolvency to the advantage of the lender. COVInsAG § 2(3) mitigates this risk by declaring that any such new loans and security granted for restructuring will not be regarded as aiding or facilitating a delay in filing for insolvency. Additionally, COVInsAG provides for the possibility for SMEs to temporarily refuse to provide payments and services until 30 June 2020 in relation to significant contracts which were entered into before 8 March 2020 if the inability to provide payments and services is due to the consequences of COVID-19 and performance would jeopardise the economic foundations of the business.

Germany's New Restructuring Process

On 1 January 2021, *Sanierungsrechtsfortentwicklungsgesetz*[236] (SanInsFoG) came into force, implementing the EU Restructuring and Insolvency Directive 2019[237] and changes recommended by a German government report on restructuring companies.[238] The core of the legislative reform is the

[234] Such transactions include performance in lieu of or on account of performance (*Leistung an Erfüllungs statt oder erfüllungshalber*), payments by a third party on the instruction of the debtor, the provision of security other than the one originally agreed if not of higher value, the shortening and relaxation of payment terms.
[235] SanInsFoG (n 222) § 12.
[236] Reorganization and Insolvency Law Development Act.
[237] Directive 2019/1023 (n 223).
[238]Bericht der Bundesregierung über die Erfahrungen mit der Anwendung des Gesetzes zur weiteren Erleichterung der Sanierung von Unternehmen (ESUG) vom 7. Dezember 2011

new Corporate Stabilisation and Restructuring Act (StaRUG).[239] StaRUG both complements the existing procedures under InsO, which already includes the possibility to restructure in self-administered proceedings on the basis of an insolvency plan and 'filled the gap' for pre-insolvency restructurings. Until StaRUG there was no legal framework in Germany for solutions outside formal insolvency proceedings. StaRUG offers companies in financial difficulties various options for pre-insolvency restructuring, which do not require universal creditor consent. This modular system includes, amongst other things, judicial proceedings for the voting on, and confirmation of, a restructuring plan with the possibility for a 'cross-class cram down' and a court-imposed ban on enforcement and realisation measures. The measures are available to companies with their Centre of Main Interest (COMI) in Germany, provided they are in a situation of imminent illiquidity (*drohende Zahlungsunfähigkeit*) and not yet under an obligation to file for insolvency.

SanInsFoG contains important amendments to the InsO. The legal definitions of insolvency events, 'imminent illiquidity' (*drohende Zahlungsunfähigkeit*) and 'over-indebtedness' (*Überschuldung*) have also been refined. In addition, SanInsFOG provides for further temporary adjustments of the InsO in response to the continuing impact of COVID-19. SanInsFoG ultimately envisages the creation of new restructuring courts to handle these new tools and to centralise proceedings and expertise.[240]

Overall, the insolvency response of Germany is not dissimilar to that of the UK. Both have focused on mitigating the consequences of COVID-19 and reducing the number of companies unduly entering

(Report the Federal Government about the experience with the application of the Act to further facilitate the restructuring of companies (ESUG) of 7 December 2011, 2582.

[239] Gesetz über den Stabilisierungs- und Restrukturierungsrahmen für Unternehmen (Unternehmensstabilisierungs- und -restrukturierungsgesetz - StaRUG).

[240] Stephan Madaus, 'A Giant Leap for German Restructuring Law? The New Draft Law for Preventive Restructuring Procedures in Germany' (*University of Oxford Faculty of Law*, 26 October 2020) <www.law.ox.ac.uk/business-law-blog/blog/2020/10/giant-leap-german-restructuring-law-new-draft-law-preventive> accessed 11 March 2021.

into insolvency proceedings. However, the process by which Germany and the UK have gone about achieving this objective is divergent due to the differing nature of each legal system. Germany suspended the director's obligation to file for insolvency, temporarily suspended directorial liability for payments made after insolvency in the ordinary course of business and prevented transactions from being challenged in subsequent insolvency proceedings where security or satisfaction of an obligation is obtained during the suspension period. The UK, on the other hand, did not suspend a director's obligation to engage with insolvency proceedings, instead adopting temporary measures suspending winding-up orders and petitions and to alleviate the risk of personal liability arising for the actions of a director who continues to trade when a company is insolvent.

Part Conclusion

The Restructuring Plan of VAA illustrates the ability of US Chapter 15 bankruptcy to adapt and respond to new insolvency processes introduced across the globe in light of COVID-19. The approach of the US was analysed in comparison to that of the UK highlighting the different approaches taken. The US has adopted expansionary fiscal and monetary policy with limited changed to its bankruptcy regime. While the UK and Germany adopted similar objectives, they embarked upon different approaches to alleviate the pressure on companies due to COVID-19. Germany suspended the obligation directors are under to file for insolvency while in the UK measures were introduced to mitigate the personal liability which would otherwise arise if directors continued to trade when a company is insolvent. In addition, both Germany and the UK have adopted new restructuring mechanisms with the German regime a competitive response to the UK's Part 26A restructuring plan. Both regimes are likely to prove to be highly utilised in the wake of COVID-19 with VAA already engaging with the UK Part 26A restructuring plan.

The future of UK Corporate Insolvency

Introduction

The rate of corporate failure in the UK prior to the COVID-19 pandemic had been rising at an alarming rate. Official figures released by the Insolvency Service in the fourth quarter of 2019 show that there were 17,196 underlying company insolvencies in 2019, a 6.8% increase on 2018, and the highest level of underlying insolvencies since 2013.[241] In recent years high-profile corporate failures have seemingly become commonplace in the UK.[242] The collapse of Carillion in January 2018, the largest corporate collapse in the country for a decade, prompted the government to conduct a consultation into its collapse.[243] This came shortly after the high-profile failure of British Home Stores[244] and was followed by the failures of British Steel in May 2019[245] and Thomas Cook in September 2019.[246] Moreover, the COVID-19 pandemic has led to further corporate failures such as Flybe in February 2020[247] and Debenhams in April 2020[248] and is likely to lead to more as the temporary support measures are removed. This part focuses on the anticipated short-term spike in insolvencies that will proceed the relaxation of temporary measures, and the framework that underpins UK corporate insolvency law in order to identify how in the future the UK insolvency regime may develop. Ultimately analysing the long-term direction of the UK insolvency regime given the recent developments contained in CIGA, which align the UK closer to the US bankruptcy regime than ever before.

---

[241] The Insolvency Service, 'Quarterly Company Insolvency Statistics, Q4 October to December 2020' (n 64).
[242] Edwin Mujih, 'Corporate governance reform and corporate failure in the UK' (2021) 42(4) Comp Law 109.
[243] ibid and Federico Mor, Lorraine Conway, Djuna Thurley and others, *Briefing Paper: The collapse of Carillion* (House of Commons Library, 14 March 2018).
[244] *Re BHS Ltd (In Administration)* [2016] EWHC 1965 (Ch), *Re SHB Realisations Ltd (formerly BHS Ltd) (In Liquidation)* [2018] EWHC 402 (Ch).
[245] *Re British Steel Ltd* [2019] EWHC 1304 (Ch).
[246] *Re Thomas Cook Group Plc* [2019] EWHC 2626 (Ch).
[247] *Re Flybe Group Plc* [2019] EWHC 631 (Ch).
[248] *Re Debenhams Plc* [2020] EWHC 1755 (Ch).

The 'Zombies' of the Pandemic

The year 2020 has seen the lowest annual rate of insolvencies for over a decade, down 4,639 companies, 27 per cent, when compared to 2019.[249] However, this is unlikely to be a true representation of the performance of companies prospering during the difficulties of COVID-19, and instead a result of the temporary measures introduced by the UK government to support companies. In this regard, it is likely that thousands of companies in the UK have become pandemic 'zombie' companies.[250] Only able to service their liabilities as a result of government support packages.

Prior to COVID-19, it was estimated that 96,000 companies (roughly 5 per cent of the total companies in the UK) could only afford to service the interest on their debt obligations rather than the principal debt itself.[251] This figure will have increased as a result of the billions of pounds which has been made available by the government to companies during the course of the pandemic through; the Business Interruption Loan Scheme, Bounce Back Loan Scheme, VAT deferrals and other support Schemes.[252] Undoubtedly this support has enabled companies that would have otherwise entered into insolvency proceedings to continue operating, creating pandemic 'zombie' companies. Ultimately when financial support measures are lifted, and the temporary measures offered through CIGA removed, the UK will be faced with a torrent of financially distressed companies engaging in insolvency proceedings. This is perhaps where the permanent measures contained in CIGA are intended to assist financially distressed companies, enabling restructuring and ultimately rescue. However, mechanisms such as the new moratorium period[253] will be of limited use to SMEs as discussed in part two, rendering it of little utility to 99 per cent of all companies in the UK.[254]

---

[249] The Insolvency Service, 'Quarterly Company Insolvency Statistics, Q4 October to December 2020' (n 64).
[250] Finch and Milman (n 4).
[251] Caroline Sumner, 'Rescue, Recovery & Renewal' (2017) 5 CRI 188.
[252] *Budget 2021: Protecting the Jobs and Livelihoods of the British People* (HM Treasury 2021).
[253] Corporate Insolvency and Governance Act 2020.
[254] Georgina Hutton and Matthew Ward, *Research Briefing: Business Statistics* (House of Commons Library 2021) 5.

This, therefore, leaves financially distressed companies to engage with the new Part 26A restructuring plan or to the pre-existing insolvency regime. This is likely to result in two main outcomes for financially distressed companies: rescue or liquidation.

The insolvency measures introduced and those already in operation do very little to secure the long-term prospects of companies. The measures introduced were at best reactionary to COVID-19 to prevent mass unemployment and the collapse of the UK economy. At worst they have created thousands of pandemic 'zombie' companies and enabled otherwise unviable companies to remain 'operational' at the expense of the UK taxpayer. The latter has been illustrated by the spike in insolvencies which has proceeded the exit of lockdowns in England.[255] Overall, the short-term holds a high number of companies interacting with the UK insolvency regime when the impact of the contraction of the UK and global economy is realised due to the phasing out of government support schemes.[256]

The future direction of the UK Insolvency regime

CIGA represents the largest change to the UK's corporate insolvency regime in more than 20 years[257] and a strengthening of the UK's rescue culture, but perhaps also a shift in the long-term direction of the regime. The need for a new restructuring procedure was proposed by the government on the basis that the British system did not have a formal debtor-in-possession procedure capable of binding all creditors through the use of 'cross-class cram down' provisions.[258] Though unacknowledged in the consultation papers the Insolvency Directive 2019[259] will undoubtably have influenced the governments

[255] *Monthly Insolvency Statistics* (The Insolvency Service 2021) 5.
[256] UK GDP in January 2021 was 9.0% lower than the level in February 2020 the last full month of 'normal' operating conditions: *Coronavirus and the impact on output in the UK economy: January 2021* (Office for National Statistics 2021).
GDP in 2020 fell by 9.9%, the largest annual fall in 300 years: 'A millennium of macroeconomic data' (*Bank of England*)
<www.bankofengland.co.uk/statistics/research-datasets> accessed 25 March 2021.
[257] UK Government, 'Major changes to insolvency law come into force' (n 78).
[258] BEIS, *Insolvency and Corporate Governance: Government response* (n 79) 63.
[259] Directive 2019/1023 (n 223).

decision. Part 26A (discussed in part two) fills this void along with making the insolvency regime more efficient, flexible and adaptable. However, this comes at the expense of transparency, fairness and accountability.[260]

Some commentators argue that the changes have been too similar to existing provisions in company and insolvency law and as a result were not needed because they add further complexity to the existing framework.[261] These commentators were probably referring to innovative practices developed by the insolvency industry to deal with companies in distress, in particular 'light touch' administration (LTA).[262] LTAs have been utilised in previous high-profile cases such as Railtrack, Metronet and Turner and Newall.[263] More recently, Debenhams announced its intention to use LTA to turn around its business,[264] with evidence that several UK retailers and restaurant chains such as Oasis & Warehouse may make use of LTAs during COVID-19.[265]

The new insolvency mechanisms introduced by CIGA aligns the UK closer than ever before with the US regime. Constituent parts of the Part 26A restructuring plan such as the 'relevant alternative'[266] are seemingly inspired by Chapter 11[267] as well as the carve outs to *ipso facto* clauses largely mirroring those found in the US regime. Despite alignment with the US bankruptcy regime, it would be incorrect to assume the UK has tried to emulate US style bankruptcy. Unlike many European countries, the UK has not used Chapter 11 as a source of inspiration for enacting and developing its restructuring

---

[260] Eugenio Vaccari, 'Corporate insolvency reforms in England: rescuing a "broken bench"? A critical analysis of light touch administrations and new restructuring plans' 2020, 31(12) ICCLR 645, 647.
[261] BEIS, *Insolvency and Corporate Governance: Government response* (n 79) 63.
[262] Insolvency Act 1986, sch B1, para 64. For discussion see, Vaccari (n 260).
[263] ibid 647.
[264] ibid.
[265] Elias Jahshan, '2020 immediate job cuts as Oasis & Warehouse files for administration' (*Retail Gazette*, 15 April 2020) <www.retailgazette.co.uk/blog/2020/04/202-immediate-job-cuts-as-oasis-warehouse-files-for-administration/> accessed 2 October 2020.
[266] Companies Act 2006, s 901G(5).
[267] 11 US Code § 1129(a)(10).

proceedings.[268] Instead, developments in UK insolvency law have come about through independent consultation. In this regard, the Insolvency Service in 2016[269] and the Department for Business, Energy and Industrial Strategy (BEIS) in 2018[270] engaged in large scale consultations and proposed recommendations to develop and improve the insolvency regime. Both consultations contained recommendations which were subsequently implemented in CIGA. Additionally, many of the recommendations implemented into UK insolvency law originate from the Cork Report, such as administration, and most recently a standalone moratorium procedure.[271]

World Bank data from 2020 ranked the US as the second-best regime in the world, with the UK at fourteenth.[272] However, when analysing the data in more detail, the UK fares better than the US in the terms of cost of insolvency, and the amount returned to creditors. On average, insolvency proceedings take one year in both the UK and the US. However, proceedings cost 6 percent of the estate in the UK compared with 10 percent in the US, and the UK regime recovers 85.4 cents in the dollar compared with 81 cents in the dollar in the US.[273] Therefore, the UK has a cheaper insolvency regime that returns more money to creditors than its US counterpart. Consequently, it is nonsensical for the UK to try and emulate the US regime. Additionally, Chapter 11 Bankruptcy has become too expensive, particularly for SMEs.[274] To adopt a US-style insolvency regime would be extremely detrimental to 99 per cent of UK companies.[275] Overall, the UK may have adopted measures that are present in the US

[268] For example, Germany, Italy, the Netherlands and Spain. For discussion see, Samantha Renssen, 'Corporate Restructuring and Corporate Dissolution of Companies in Financial Distress: Ensuring Creditor Protection. A Comparison of the US, UK and Dutch Models' (2017) 26 Int'l Insolvency Rev 204 205.
[269] Insolvency Service, *A review of the Corporate Insolvency Framework* (n 41).
[270] BEIS, *Insolvency and Corporate Governance* (n 79).
[271] Cork, *Report of the Review Committee* (n 2).
[272] 'Doing Business, Resolving Insolvency' (*World Bank Group*) <https://archive.doingbusiness.org/en/data/exploretopics/resolving-insolvency#> accessed 25 March 2021.
[273] ibid.
[274] *Commission to Study the Reform of Chapter 11 2012-2014: Final Report and Recommendations* (American Bankruptcy Institute 2014).
[275] Hutton and Ward (n 254).

bankruptcy regime. However, this does not mean that the UK intends to emulate the US regime in the future.

The UK insolvency regime is highly sophisticated, with viable companies able to make use of a plethora of rescue procedures.[276] This sophistication and flexibility[277] enables insolvency practitioners to decide whether to rescue the company or its business or to liquidate the assets when rescue is not reasonably practicable and would lead to a higher return to creditors.[278] The future direction of the UK regime is ultimately not known due to the large periods of time between developments. What is known is that the future developments will be based on the recommendations and findings of consultations which identify weaknesses in the present day regime.

Part Conclusion

Overall, the introduction of temporary support measures has created a large number of pandemic 'zombie' companies. In the short-term, the UK insolvency regime has been challenged by the volume of companies who are have engaged with the regime upon the easing of these measures. The greatest change to the UK insolvency regime for 20 years was a reaction to COVID-19 breaking the historical trend of stagnated development. It is therefore likely that significant change will not be seen again for a long time.

The changes introduced by CIGA bring the UK insolvency regime more in line with its US counterpart. That said, the UK has not sought to imitate the US regime. Such imitation would be detrimental to UK companies. The UK regime has been developed over a number of decades and CIGA represents another development which ultimately strengthens rescue culture in the UK and further entrenches the flexibility and adaptability of the regime. Future development and direction of UK insolvency law will depend on subsequent consultations by the UK government and Insolvency Service.

---

[276] Okoli (n 74) 64.
[277] Vaccari (n 260) 655.
[278] ibid.

Conclusion

The UK insolvency regime has developed from a confused tangle of statutes in the nineteenth century to a highly sophisticated, efficient, and flexible system. The Cork Report remains central to the development of the UK regime over forty years later with measures implemented by CIGA first being recommended by Cork and subsequent consultations. CIGA has further strengthened the UK insolvency regime introducing a new standalone moratorium period, broadening the ban on *ipso facto* clauses and establishing a cross-class cramdown procedure through the new Part 26A restructuring plan. The utility of the Part 26A restructuring plan has been observed through the landmark cases *Re Virgin Atlantic Airways*[279] and *Re Deepocean 1 Ltd.*[280] However, the strengthening and development of the regime has potentially been hindered by the partial reintroduction of Crown preference which has the potential to erode the utility of CVAs.

The fall in demand resulting from the introduction of travel restrictions in response to COVID-19 has contributed to many aviation companies' financial difficulties. The difficulties of VAA were analysed in detail, in doing so discussing the Part 26A restructuring plan the company has engaged with and the subsequent US Chapter 15 bankruptcy protection granted over the company's assets. The cross-border nature of *Re Virgin Atlantic Airways*[281] allowed comparison to be drawn between the US and UK insolvency response to COVID-19, wherefore it was identified the US has adopted an expansionary monetary and fiscal policy with minimal regime change. This contrasts with the response of the UK and Germany who have both implemented new restructuring mechanisms and pursued temporary change to reduce the number of companies interacting with their insolvency regimes and encourage companies to continue trading.

The support measures introduced in response to COVID-19 resulted in the lowest rate of insolvency in the UK for over a decade.

[279] *Re Virgin Atlantic Airways Ltd* (n 6).
[280] *Re Deepocean 1 UK Ltd* (n 7).
[281] *Re Virgin Atlantic Airways Ltd* (n 6).

As measures are eased and companies exposed to the economic contraction caused by COVID-19 a large number of the pandemic 'zombie' companies have engaged with the UK insolvency regime. CIGA aligns the UK closer to the US bankruptcy regime than ever before, however this does not represent a shift in the UK regime to imitate that of the US. Subsequent development of the UK insolvency regime will be in response to future consultations and reviews, and not by replicating the US regime.

Overall, COVD-19 has resulted in the introduction of permanent and temporary changes to the UK insolvency regime, changes which have been emulated to a large extent by Germany but rejected by the US. The true impact of COVID-19 is yet to be felt by the UK insolvency regime as companies are sheltered by protection schemes, only once they are entirely lifted will the true impact of COVID-19 be realised.

Subsequent Cases

Re Gategroup Guarantee Ltd [2021] EWHC 304 (Ch)
Re Port Finance Investment Ltd [2021] EWHC 378 (Ch)
Re Smile Telecoms Holdings Ltd [2021] EWHC 395 (Ch)
Re Virgin Active Holdings Ltd [2021] EWHC 814 (Ch)
Re All Scheme Ltd [2021] EWHC 1401 (Ch)
Re National Car Parks Ltd [2021] EWHC 1653 (Ch)
Re Hurricane Energy Plc [2021] EWHC 1759 (Ch)
Re Amicus Finance Plc [2021] EWHC 2255 (Ch)
Re CFG Investments SAC [2021] EWHC 2780 (Ch)
Re MAB Leasing Ltd [2021] EWHC 152 (Ch)

The Meaning and Role of 'Without Legal Ground' in the Scots Law of Unjustified Enrichment

RAIYAN CHOWDHURY*

Abstract

The 'Without Legal Ground' concept is an aspect of the law of unjustified enrichment in Scotland. Though it has a long history, it was only fully cemented into Scots Law with two major cases in the 1990s, Shilliday v Smith *1998 SC 725 and* Dollar Land v Cumbernauld *1998 SC (HL) 90. However, these cases fail to confirm the meaning and role of the concept, and more than 20 years later the law has not advanced far. This article seeks to ascertain both the meaning and role of the concept. It concludes that its meaning must differ according to the manner of the enrichment. Further, the proper role of the concept is as a subsidiary principle to develop the law. However, it also argues that the value of the concept differs according to the manner of the enrichment. In deliberate transfer cases, it can be a useful tool in novel situations. However, in other types of cases, it cannot be relied upon in this manner and instead is merely normative justification to develop the law.*

Keywords: unjustified enrichment, Scots law, private law, law of obligations

Introduction

'In general terms it may be said that the remedy is available where the enrichment lacks a legal ground to justify the retention of the benefit.'[1]

It was with the above statement from Lord Hope that the notion of 'without legal ground' was cemented into the Scots law of unjustified enrichment. More than 20 years since the judgment of the House of Lords in *Dollar Land*, it remains unclear what exactly the

* DPLP Student, University of Aberdeen.

[1] *Dollar Land (Cumbernauld) Ltd v CIN Properties Ltd* 1998 SC (HL) 90, 98.

concept means, and what role it has (or should have) in the law. Is it a guiding principle, a component of a general action, or a meaningless label for this area of the law? The questions of what the concept means and what its role is/should be are tied together, as suggesting a broad meaning attaches a greater role to the concept. This article will begin with an overview of the development of the 'without legal ground' concept, both in Scots law and the broader Civilian Tradition. It will then explore the possible roles of the concept, first as a general action of enrichment, second as a subsidiary principle to develop the law. Consideration will also be given as to whether it is unable to serve any meaningful role and should be discarded. It will be necessary to look to other jurisdictions for guidance, chiefly Germany and South Africa. It will be shown that 'without legal ground' has a role as a subsidiary principle, but its meaning and value differ according to the manner of enrichment.

The Development of 'Without Legal Ground'

The immediate history of the concept of 'without legal ground' begins with the Inner House judgment in *Dollar Land*,[2] in which Lord Cullen said:

> A person may be said to have received unjustified enrichment at another's expense when he has obtained a benefit from his actings or expenditure, without there being a legal ground which would justify him in retaining that benefit.[3]

This statement was wholly adopted by Lord President[4] Rodger in *Shilliday v Smith*.[5] The House of Lords decision in *Dollar Land* came after *Shilliday*, and it was here that Lord Hope laid down the general principle quoted in the introduction. He also laid out some further requirements for a claim (discussed below).[6]

---

[2] *Dollar Land (Cumbernauld) Ltd v CIN Properties Ltd* 1996 SC 331 (IH).
[3] ibid, 348-349.
[4] As he then was.
[5] 1998 SC 725, 727 (IH).
[6] *Dollar Land* (n 1) 99.

While *Dollar Land* and *Shilliday* are rightly regarded as part of the Scottish 'Enrichment Revolution'[7], as Evans-Jones highlights, the general enrichment principle 'is not wholly a new beginning'.[8] Evans-Jones argues that the concept of 'without legal ground' functioned as a broad principle within the *condictio* group of claims.[9] Stair, in his *Institutions*, stated that:

> The duty of restitution extendeth to those things, *quae cadunt in non causam*, which coming warrantably to our hands… yet if the cause cease by which they become ours, there superveneth the obligation of restitution of them'.[10]

This passage is cited as authority for the existence of the 'without legal ground' approach pre-*Dollar Land*.[11] The *quae cadunt in non causam* maxim has been argued to represent the 'without legal ground' concept,[12] though Trayner gives it a slightly different definition.[13] Perhaps more representative of the unifying principle of the *condictiones* is the notion of an absence of cause for the transfer.[14] Du Plessis argues that 'cause' was not regarded as a general requirement of liability but rather is used to describe a specific situation where a claim would be available.[15] Indeed, Evans-Jones highlights that the principle 'had little impact on the law until recently'.[16] In any case, the impact of *Dollar Land* is to elevate the

---

[7] Niall R Whitty, 'The Scottish Enrichment Revolution' (2001) 6 SLPQ 167.
[8] Robin Evans-Jones, 'Thinking about Principles and Actions: Unjustified Enrichment in Scots and South African Law', in Douglas Bain, Roderick R M Paisley, Andrew R Simpson and others (eds) *Northern Lights: Essays in Private Law in Memory of David Carey Miller* (Aberdeen UP 2018) 336.
[9] ibid 324.
[10] Stair, *Institutions*, I, 7, 7.
[11] Robin-Evans Jones, *Unjustified Enrichment*, vol 2 (W Green 2013) para 3.31.
[12] ibid, para 2.31.
[13] John Trayner, *Latin Phrases and Maxims* (Edinburgh 1861) 271.
[14] West finds the *quae cadunt* maxim to be equivalent to *sine causa*. See Euan West, 'The Scots Law of Unjustified Enrichment: The Adequacy of a Pure 'Without Legal Ground' Approach' (Masters Thesis, University of Aberdeen 2015) 40.
[15] Jacques Du Plessis, 'Towards a Rational Structure of Liability for Unjustified Enrichment: Thoughts from Two Mixed Jurisdictions' (2005) 122 SALJ 142, 148.
[16] Robin Evans-Jones, 'Thinking about Some Scots Law: Lord Rodger and Unjustified Enrichment' in Andrew Burrows, David Johnston and Reinhard Zimmermann (eds) *Judge and Jurist: Essays in Memory of Lord Rodger of Earlsferry* (OUP 2013) 431, 443.

notion of 'without legal ground' to cover all cases of unjustified enrichment, not just the *condictiones*.[17] Though Lord President Rodger conceded that the principle may not cover all cases,[18] no such qualification was found in the judgement of Lord Hope in *Dollar Land*, or the subsequent English case of *Kleinwort Benson v Lincoln City Council*.[19] Thus, what the 'enrichment revolution' has brought about is a general principle of unjustified enrichment, centred on the notion of 'without legal ground'. The question remains, what role does this principle play in actual claims under the law of unjustified enrichment? Further, what does 'without legal ground' mean in practice? As mentioned above, these questions are tied together.

It is important to make brief mention of the Roman and Civilian roots that the 'without legal ground' concept rests on. In Roman Law, there were multiple *condictio* claims covering different situations.[20] West argues that the *condictiones* provided the 'building blocks' of a general principle of 'retention *sine causa*' (without cause), though the principle was not at that time in existence.[21] Going forward to the Civilian tradition, Du Plessis highlights how the jurist Grotius formulated the general principle of deriving benefit without *legal title*.[22] Two centuries later, the jurist Savigny recognised as a common characteristic of the Roman *condictiones* that there was an unjustified transfer of wealth from one person to another, and that said transfer was unjustified due to the absence of a valid legal ground (*causa*).[23] This will be relevant for our discussion of the possible general enrichment action.

General Enrichment Action

One possible role that could be assigned to the 'without legal ground' concept is that it is part of a general enrichment action, along with the other requirements laid out by Lord Hope in *Dollar Land* (mentioned

[17] Robin Evans-Jones, *Unjustified Enrichment*, vol 1 (W Green 2003) 34.
[18] *Shilliday* (n 5) 727.
[19] [1999] 2 AC 349 (HL), 409.
[20] Eg, D. 12.6 (*condictio indebiti*) and D. 12.7 (*condictio sine causa*).
[21] West (n 14) 37.
[22] Du Plessis (n 15) 144.
[23] Reinhard Zimmermann and Jacques du Plessis, 'Basic Features of the German Law of Unjustified Enrichment' (1994) 2 RLR 14, 17.

above). It is possible to take from *Dollar Land* a 'four point test'[24] for enrichment liability:

(1) The defender is enriched.
(2) At the expense of the pursuer.
(3) The enrichment is without legal ground to justify retention of the benefit.
(4) It is equitable to compel the defender to redress the enrichment.[25]

The 'without legal ground' concept is the most important part of this test; it explains why the enrichment was unjustified. It also requires the most explanation at an abstract level.

It seems that the Court of Session has in fact used the 'without legal ground' principle as general enrichment action, for example in *McGraddie v McGraddie*.[26] The consequence of a general action is the abandonment of older nominate types of claims that have previously been pled before courts.[27] This idea was embraced by Sheriff Way in *Esposito v Barile*,[28] where he stated that it is unnecessary to 'shoehorn the facts into a particular style of Roman sandal'.[29] Is this desirable? Hogg, in a pros and cons analysis, presents the following as the 'strongest argument' in favour of a general action. Given that all enrichment claims protect a single (restitutionary) interest, this perhaps naturally suggests a unitary action (like in Delict).[30] Further, given that there is a 'golden opportunity' to reform the law, the reform should not be made more complicated than is necessary.[31] Indeed, the perceived simplicity of the general action is an attractive quality. The law should be simple to understand and utilise where possible. The

[24] Niall Whitty, 'Transco plc v Glasgow City Council: developing enrichment law after Shilliday' (2006) 10(1) Edin LR 113, 116.
[25] *Dollar Land* (n 1) 99.
[26] [2012] CSIH 23, [54], [55]. The case was overturned on unrelated grounds by the Supreme Court ([2015] UKSC 1).
[27] For example, the *condictio causa data causa non secuta* (see *Shilliday* (n 5)).
[28] 2011 Fam LR 67 (Sh Ct).
[29] ibid [17].
[30] Martin Hogg, 'Unjustified Enrichment in Scots Law Twenty Years On: Where Now' (2006) 14 RLR 1, 7.
[31] ibid 7.

existence of a (*prima facie*) general enrichment action in German law[32] lends further support to the desirability of such an action in Scots law. However, there are compelling arguments against a general action, and these will be demonstrated first by exposing some issues that the 'without legal ground' concept has created in the case law.

The main case that shows the problems created by a potential general enrichment action is *Virdee v Stewart*.[33] The case concerned an agreement between two parties, whereby the pursuer would build a house on the defender's land which she (the pursuer) and her family would then use from time to time.[34] The agreement broke down and the pursuer, having been excluded from the house, sought a claim in Unjustified Enrichment. If we turn immediately to the 'without legal ground' concept there is a problem. It is unclear what exactly 'without legal ground' should mean in this context. Lady Smith stated that: 'from the outset, the defender had no legal right or entitlement to have the pursuer build anything on his land or to confer any benefit on him by way of taking any action on or relating to his land'.[35] Accordingly, the defender had been unjustifiably enriched from the completion of the house (in 1994), and therefore the claim (having been made in 2010) had prescribed.[36] This decision seems flawed. How could it be that the enrichment was unjustified when the house was built, given that there was a (non-contractual) mutual understanding concerning the house between pursuer and defender? The judgment seems to focus on the strict legal entitlement of the defender to acquiring the house, inferring from the lack of which that there was no legal ground.[37] Further, Evans-Jones argues that Lady Smith proceeded on the basis of the 'building by mistake' paradigm case, while this was clearly a case of building *by agreement*. [38] In this regard, the correct basis for the claim would be a *condictio*, perhaps the *condictio causa data*

[32] Bürgerliches Gesetzbuch (BGB) §812(1).
[33] [2011] CSOH 50.
[34] ibid [3], [4].
[35] ibid [24].
[36] ibid [12], [26].
[37] Robin Evans-Jones, 'Scotland' (2011) 19 RLR 241, 243.
[38] ibid 244.

causa non secuta (*condictio c.d.)*[39] or the *condictio ob causam finitam*.[40] It follows that the enrichment was 'without legal ground' (and therefore unjustified) when the agreement broke down, and had not prescribed.[41] Thus, we can see how using the 'without legal ground' concept as a general enrichment *action* can lead to inappropriate results. In *Thomson v Mooney*[42] the 'without legal ground' concept was applied as a general action in a similarly inappropriate way. The case concerned transfers in the contemplation of marriage (the paradigm case of the *condictio c.d.*[43]) but Lord Glennie held that these transfers were gratuitous.[44] Therefore, there was no basis for retention from the moment of receipt and the claim had therefore prescribed.[45]

Thankfully, *Thomson* was overturned on appeal.[46] The Inner House relied on the specific claim of the *condictio c.d.* and relevant authority.[47] This leads us to the critical point. If we are to have a general enrichment action arising from the 'without legal ground' concept, then this concept needs to be understood in light of existing types of claims. Otherwise, we are left with three meaningless words and the courts would have to stumble around in the dark whenever an enrichment claim is brought before them. The experience of the general enrichment action in German law is telling in this regard. In Germany, the nominate *condictio* claims are specifically mentioned alongside this action in the BGB (the German Civil Code).[48] Further, as Zimmerman and Du Plessis highlight, it is difficult to interpret the terms 'at the expense of' and 'without legal ground'[49] without an 'uncontrollable extension of liability'.[50] The major breakthrough for German enrichment law came in the work of Walter Wilburg and

[39] Though this was rejected by Lady Smith, see *Virdee* (n 33) [17], [26].
[40] See Martin Hogg, 'Unjustified enrichment claims: when does the prescriptive clock begin to run?' (2013) 17(3) Edin LR 405, 407.
[41] ibid 407.
[42] [2012] CSOH 177.
[43] Stair, *Institutions*, I, 7, 7.
[44] *Thomson* (n 42) [12].
[45] ibid [12], [18]. For criticism, see Hogg (n 40) 407-410.
[46] [2013] CSIH 115.
[47] Stair, *Institutions*, I, 7, 7 and *Shilliday* (n 5).
[48] Zimmerman and Du Plessis (n 23) 18-19. These include the *condictio c.d.* in BGB §812(I)(2), and c*ondictio ob turpem vel iniustam* causam in BGB §817(1).
[49] BGB §812(I).
[50] Zimmerman and Du Plessis (n 23) 24.

(later) Ernst von Caemmerer. Wilburg recognised the fundamental distinction between enrichment claims that arose from a performance (*Leistung*) and claims that arose in any other manner.[51] The importance of this distinction is that the general action in §812(I) BGB must be understood differently according to the manner of the enrichment. For example, the 'at the expense of' requirement only relates to the 'other' category of claims.[52] More relevant for our purposes is the contention that the Wilburg/von Caemmerer taxonomy casts light on what 'without legal ground' *means*. In the deliberate transfer cases, the concept focuses on whether the purpose of the transfer has been achieved, and if it has not then the enrichment is without legal ground.[53] It is fair to say that this approach to unjustified enrichment has wide support in academia and in German courts.[54] The Wilburg/von Caemmerer approach of assigning multiple meanings to 'without legal ground' makes the concept far more intelligible.[55]

The German approach is one that enjoys academic support in Scotland also. According to Evans-Jones, the 'without legal ground' principle 'should be understood differently according to the different groups of cases to which it is applied'.[56] Accordingly, the principle would focus on failure of purpose in the cases concerning the *condictiones* (deliberate conferral/transfer).[57] Of course, even with the concession that the principle must be understood differently according to different groups of claims, one can still argue that there should be a general action. Per Hogg, questions of taxonomy can be separated from whether there should be a general enrichment action.[58] However, if the starting point is the manner of enrichment, this weakens the 'generality' of the action. Further, Evans-Jones argues

[51] ibid 25.
[52] ibid 25.
[53] ibid 26. This is the 'subjective approach', see: Jacques du Plessis, 'Labels and Meaning: Unjust Factors and Failure of Purpose as Reasons for Reversing Enrichment by Transfer' (2014) 18 Edin LR 416, 431-432.
[54] See Zimmerman and Du Plessis (n 23).
[55] West (n 14) 58.
[56] Evans Jones, vol 2 (n 11) 3.35. See also West (n 14) 81, and Martin Hogg, 'Continued uncertainty in the analysis of Unjustified Enrichment' (2013) 15 SLT 111, 112.
[57] Evans-Jones, vol 1 (n 17) 45. West argues that the principle should be understood in both a subjective and objective way, see West (n 14) 80.
[58] Hogg (n 30) 8.

that approaching claims from a single abstract ground can elide the concrete conditions for established claims.[59] The law should not look towards an 'unknowable justice in the sky'.[60] Indeed, established claims in Unjustified Enrichment have specific requirements, such as the necessity of error in the *condictio indebiti*.[61] These requirements control the scope of the law, and a general action that loses sight of them may allow unbounded extension of liability. Also, it obscures the different policy considerations that underlines the different established claims. It is submitted therefore that the 'without legal ground' principle should not be elevated to the role of a general enrichment action. As Hogg states:

> A pursuer should be asked to identify the mode by which the defender acquired the enrichment, and to make out the specific requirements of the case applicable to such mode.[62]

'Without Legal Ground' as a Subsidiary Principle

According to Evans-Jones, the general enrichment principle must be subsidiary to the established causes of action.[63] While all established claims may be rationalised *ex post facto* as emanating from the principle,[64] those established categories remain. The main function of the principle is to provide for development of the law. It allows new fact situations that do not fit into the established claims to be brought within the law of Unjustified Enrichment but provides stability by requiring that these new situations be sufficiently analogous to existing claims.[65] An example of this is the *condictio sine causa*, which exists within the *condictio* group of claims as a general action to recover what is retained without a legal basis. Per Evans-Jones, the claim arises when there is purpose for the transfer that is identical/sufficiently analogous with a nominate cause of action (e.g.,

---

[59] Evans-Jones, vol 2 (n 11) para 3.39.
[60] Peter Birks, *An Introduction to the Law of Restitution* (2nd edn, OUP 1989) 19.
[61] *Morgan Guaranty Trust Company of New York v Lothian Regional Council* 1995 SC 151, 165 (IH).
[62] Hogg, 'Continued Uncertainty' (n 56) 112.
[63] Evans-Jones, vol 2 (n 11) para 3.37.
[64] See diagram at ibid para 3.02.
[65] ibid 3.14.

the *condictio indebiti*) and the purpose has failed.[66] It can be used to recover a gift that failed,[67] or a transfer made in *doubt* (rather than error) as to legal liability.[68]

The position of Evans-Jones is supported by other jurisdictions. In Germany, an unenacted redraft of the Unjustified Enrichment codification contains a subsidiary general action, which can be used for fact specific situations outwith but analogous to the existing claims.[69] The Supreme Court of Appeal of South Africa in *McCarthy Retail (Pty) Ltd v Shortdistance Carriers CC*[70] seemingly left the door open for the recognition of a general enrichment action.[71] Seizing on this, Visser argues that a subsidiary action is the only 'reasonable option', and that the South African courts would not throw away all existing enrichment actions in favour of an 'amorphous' general action. [72]

Thus, it seems that there is a role for the 'without legal ground' principle. It can be subsidiary principle beneath the established claims. Whenever a novel situation comes before the courts, the principle can be used to dictate whether the claim is sufficiently analogous to existing claims, and therefore whether it can be accepted into the law of Unjustified Enrichment. This helps avoid unjust results in cases which do not fully conform to the requirements of existing claims. At the same time, the established claims, which are well developed, in Scots law, are retained. However, there are some issues with this idea, which will be explored below.

---

[66] Evans-Jones, vol 1 (n 17) 173.
[67] A *condictio donandi causa*. See e.g. *Findlay v Monro* (1698) Mor 1767 (Court of Session).
[68] Eg *Balfour v Smith & Logan* (1877) 4R. 454 (CSIH). See Evans-Jones, vol 1 (n 34) 183-184.
[69] An English translation is available in Reinhard Zimmermann, 'Unjustified Enrichment: The Modern Civilian Approach' (1995) 15 OJLS 403, 429.
[70] [2001] ZASCA 14, [2001] 3 All SA 236 (A).
[71] ibid [8]-[10].
[72] Daniel Visser, 'The Potential Role of a General Enrichment Action' (2009) 20 Stellenbosch L Rev 454, 457-458.

A Mere Label?

There is some force in the argument that the 'without legal ground' principle is merely a label that the law applies in certain situations, but which has no analytical value. This cannot be sustained as regards to the *condictiones*, as the translation of 'without legal ground' to 'failure of purpose' does give a useful concept to both unify the *condictiones* and to allow for development beyond the nominate causes of action. It gives the reason why there is enrichment liability in 'transfer' cases.[73] However, the same cannot be said for the 'other' cases, imposition[74] and interference.[75] Scott argues that:

> [n]o general concept can be formulated which captures the meaning of "legal ground" across all these cases, in the sense that its absence furnishes the central reason for restitution.[76]

Of course, we have already concluded that the legal ground concept must be understood differently according to the manner of the enrichment. However, while 'failure of purpose' does provide a central reason for restitution in *condictio* cases, it is debatable whether there can be an understanding of the concept in the imposition cases and the interference cases which does the same amount of normative work.

For example, in the context of imposition, Evans-Jones argues that:

> a benefit is retained without a legal ground if it was imposed upon the defender according to the paradigm cause of action of the good faith possessor/builder on another's land, or; the cause of action arose from discharge of another's money debt.[77]

---

[73] See Jacques du Plessis, 'Labels and Meaning: Unjust Factors and Failure of Purpose as Reasons for Reversing Enrichment by Transfer' (2014) 18 Edin LR 416, 432.
[74] See Evans-Jones, vol 2 (n 11) paras 4.01-4.109.
[75] ibid paras 5.01-5.117.
[76] Helen Scott, 'Rationalising the South African Law of Enrichment' (2014) 18 Edin LR 433, 436.
[77] Evans-Jones, vol 2 (n 11) para 3.36.

Putting the issue of payment of another's debt to one side, this definition proceeds directly from the paradigm case and tells us little about how the law may develop analogously from there. Similarly, for interference cases, Evans-Jones argues that: 'a benefit is retained without a legal ground if it was acquired by interference with the pursuer's patrimonial rights as that is technically defined'.[78] Evans-Jones recognises that both these definitions rely on the paradigm cases, stating that they reflect the '*core* understandings.'[79] It can be questioned then why reference to a general principle is needed, when the law is can merely develop analogously without any use for the principle. In a deliberate transfer case, a judge confronted with a novel situation may question whether there has been a failure of purpose (so the benefit has been retained without legal ground) and then consider how analogous this situation is with the existing *condictio* claims. In 'other' cases, whether imposition or interference, the line between overarching principle and paradigm case is not particularly clear. Thus, the concept of 'without legal ground' provides little to help judges develop the law. In fact, it is perhaps more likely to be a distraction, an invitation to forgo past development in favour of using the principle as a general enrichment action. This reflects the fact that the 'without legal ground' principally historically reflected only the *condictiones*.

It is therefore arguable that the 'without legal ground' principle is merely a label beyond the *condictiones*. It is simply a statement that unifies the law of unjustified enrichment under a single concept. However, looking at the South African experience reveals a counter argument to this proposition. In the 1966 case of *Nortje v Pool*,[80] the Appellate Division of South Africa refused to recognise a general enrichment action. Visser argues that the judgment was 'disastrous' for the South African enrichment law, as it brought about stagnation.[81] Decisions following *Nortje* failed develop or refine the law, and the fundamental shape of enrichment law was '"frozen" in a state of

[78] ibid.
[79] ibid.
[80] 1966 (3) SA 96 (A).
[81] Daniel Visser, 'Unjustified Enrichment' in Reinhard Zimmerman and Daniel Visser (eds), *Southern Cross: Civil Law and Common Law in South Africa* (OUP 1996) 554.

suspended animation'.[82] This might lend some insight into why an overarching principle based on 'without legal ground' is considered useful. While it does not provide analytical value in cases outside the *condictiones*, it provides normative justification for analogous development. Without it, the law may be cursed to stay within rigid categories of causes of action, not permitting any new causes of action even when it may be just to do so. This, it is submitted, is the real value of 'without legal ground' beyond the *condictiones*. While still a subsidiary principle, the value of the concept is lessened as it goes from being an analytical tool to simply a normative justification for development.

Of course, it may be argued that whatever normative value 'without legal ground' may have, this value has thus far failed to materialise in Scots law. The law of unjustified enrichment has failed to develop meaningfully since *Dollar Land* in many aspects. Per MacQueen, '[t]here is still a long way to go before the "enrichment revolution" triggered by the great trilogy of 1990s cases… can be regarded as fully bedded in'.[83] Nevertheless, this does not mean that in the future the 'without legal ground' principle cannot be used in the manner discussed to develop the law, providing that the judiciary show willingness to rationalise and progress the law in this area.

Conclusion

Given that it has been more than 20 years since the judgement in *Dollar Land*, it is somewhat disheartening that Scots law has still not clearly defined the role and meaning of the 'without legal ground' principle. The above discussion has shown some of the pitfalls with using the principle as a general enrichment action, as some judges have done. Instead, it is submitted that the 'without legal ground' concept should be a subsidiary principle that is used to develop the law in new situations. Further, its meaning should differ depending on the manner of the enrichment in question. While the principle is not a mere label, it should also be recognised that its value differs depending on the manner of the enrichment. While in the *condictio*

[82] ibid 554.

[83] Hector MacQueen, 'Cohabitants, unjustified enrichment, contract and subsidiarity: Pert v McCaffrey' (2020) 165 Fam LB 4, 5-6.

cases it actively assists the development of the law, in other categories it simply encourages development.

A Proposal for Physician-Assisted Dying in the UK

HELENE HAUSKELLER*

Abstract

Although people are living for longer, many spend the final period of their lives in pain and suffering. Some would rather die than continue living with crippling, fatal illness. This has fuelled the debate around the right to die and the morality and legality of physician-assisted dying (PAD). This article will offer a proposal for effective decriminalisation of PAD in the UK. It will discuss the key issues that would need to be addressed before setting out the eligibility requirements, threshold conditions and safeguards required against the inherent risks of legalised PAD. Finally, this article will consider the practical aspects of PAD and propose the establishment of an arms-length authority.

Keywords: assisted dying, medical law

Introduction

Modern advances in medical technology and intervention have led to an increase in the average human life span by increasing the chances of survival as well as prolonging the amount of time patients can live with a severe illness or disease.[1] In addition, modern medicine has enabled life-sustaining treatments that keep patients alive for a prolonged period, often in an unconscious state and without any prospect of recovery, when a patient is already brain dead.[2] These conditions have motivated ethicists, physicians, and patients to claim that individuals who experience incurable and irremediable suffering

* LLB, University of Aberdeen, 2021. The author is currently working at a law firm in Devon while applying for training contracts and is looking forward to becoming a qualified solicitor.

[1] James C Riley, *Rising Life Expectancy: A Global History* (CUP 2001).

[2] Helga Kuhse, Udo Schüklenk and Peter Singer, *Bioethics: An Anthology* (3rd edn, Wiley Blackwell 2015) 367.

should be granted a human right to die in an autonomous and dignified death, at a time and place of their own choosing. Currently UK law does not grant patients such a right.

As the law stands, assisting someone to die is a crime of its own in the UK.[3] If it cannot be proven that the person wanted to die voluntarily, assisting someone to die can be prosecuted as murder or manslaughter. Positive legalisation of physician-assisted dying (PAD) would require the repeal of, thereby legalising all assistance to commit suicide, and in turn making it very difficult to reduce the risk posed to vulnerable people. An alternative approach, for which there is already precedent in the UK in regard to abortion law, would be to instead decriminalise PAD. This would allow access to PAD to be restricted to a defined group of patients and enable the implementation of necessary safeguards. It is also important to note that health is a devolved matter.[4] Therefore, in order for PAD to be decriminalised across the UK, the devolved administrations would each have to legislate to that effect.[5]

Although several draft bills have been introduced into the British parliament with the aim of granting access to PAD to a defined group of patients, so far none of them have been enacted.[6] The most recent is the Assisted Dying Bill[7] which, at the time of writing, is awaiting its second reding in the House of Lords.[8] The Assisted Dying

[3] In England and Wales, under the Suicide Act 1961, s 2(1)(a). In Northern Ireland, under the Criminal Justice Act (Northern Ireland) 1966, s 13. The position in Scotland is less clear, and generally unsatisfactory: James Chalmers, 'Assisted suicide: jurisdiction and discretion' (2010) 14 Edinburgh Law Review 295.

[4] Scotland Act 1998, s 29, Government of Wales Act 2006, s 94 and Northern Ireland Act 1998, s 6.

[5] For more information see, for example, 'Devolution' (*Northern Ireland Assembly*) <https://education.niassembly.gov.uk/post_16/snapshots_of_devolution/gfa/devolution> accessed 3 May 2022 or 'Debate Call After Euthanasia Poll' (*BBC News*, 9 September 2004) <http://news.bbc.co.uk/1/hi/scotland/3640438.stm> accessed 3 May 2022.

[6] Assisted Dying for the Terminally Ill HL Bill (2003-04) 17; Assisted Dying HL Bill (2014-15) 6; and Assisted Dying (No 2) HC Bill (2015-16) [7]; among others.

[7] Assisted Dying HL Bill (2019-21) 69.

[8] Carolyn Doyle, 'Assisted Dying Bill - My Reflections and Thoughts' (*Royal College of Nursing*, 26 February 2020) <www.rcn.org.uk/news-and-events/blogs/assisted-dying-bill-my-reflections-and-thoughts> accessed 3 May 2022.

Bill sets out some reasonable conditions and safeguards for decriminalising PAD. However, it omits other safeguards which are essential for comprehensive and effective implementation. For example, it only decriminalises one form of medical assistance to die, namely physician-assisted suicide (PAS), it limits access to patients with a terminal illness and that it does not include safeguards regarding initial request and level of suffering.

This article offers a different proposal for decriminalising PAD in the UK. It discusses the specific conditions and safeguards needed for preventing misuse of PAD, protecting patients and physicians, and ensuring consistency with other individual and human rights already granted in the UK. This proposal is based on findings from an examination of legal literature, including national and foreign legislation and case law, scholarly articles, medical and bioethical works, and wider audience articles.

The first chapter examines the shortcomings of the current law and why it is inadequate. The second chapter discusses a variety of questions which will demand careful consideration by lawmakers to ensure that any reform is just and effective. These include questions around eligibility criteria and threshold conditions. The third chapter considers the safeguards required to afford the necessary protections to patients, physicians, and society. The final chapter examines the three options for practical application of the proposed conditions and safeguards, including a proposal for the establishment of an arms-length statutory authority.

Forms of Medical Assistance to Die

There are two forms of medical assistance to end someone's life voluntarily. The first is PAS, which is currently illegal and defined in the England and Wales as either 'encouraging or assisting' a patient in committing suicide,[9] and is usually in the form of prescribing a patient with a life-ending drug. PAS carries a sentence of up to fourteen years imprisonment.[10] In Scotland, such an act could lead to prosecution for murder or culpable homicide, and punishment of up

[9] Suicide Act (n 3) s 2(1)(a).
[10] ibid s 2(1C). Also see Criminal Justice Act (Northern Ireland) 1966, s 13.

to life imprisonment.[11] The second form is active voluntary euthanasia (AVE). In this article we will focus on AVE carried out by physicians, which is actively causing the death of a patient, with their consent, usually by injecting them with a lethal drug.[12] The English and Welsh courts regard AVE as either murder or manslaughter with a maximum sentence of life imprisonment.[13] The difference between PAS and AVE is whether the patient carries out the final act themselves or whether the physician carries it out for them.[14]

The Assisted Dying Bill proposes the decriminalisation PAS only,[15] as has been done in Switzerland, Germany, and the US State of Oregon.[16] This approach would remove the doubt surrounding the patient's true intentions and wishes, which would inevitably remain where the final act is carried out by someone else.[17] However, for comprehensive and effective decriminalisation the right to die should be afforded to patients who can carry out the final act themselves as well as those who would require assistance,[18] as has been done in the Netherlands and Canada.[19] This would be especially significant for patients with physically or psychologically deteriorating conditions who have lost or are in danger of losing the physical and mental ability to participate in PAS. This article proposes the decriminalisation of

[11] 'Dignity in Dying: 'The Law'' (*Dignity in Dying*) <www.dignityindying.org.uk/assisted-dying/the-law/> accessed 3 May 2022.
[12] 'Euthanasia and Assisted Suicide' (*NHS*) <www.nhs.uk/conditions/euthanasia-and-assisted-suicide/> accessed 3 May 2022.
[13] *R v Inglis* [2010] EWCA Crim 2637 [37]: 'the law of murder does not distinguish between murder committed for malevolent reasons and murder motivated by familial love. Subject to well established partial defences, like provocation or diminished responsibility, mercy killing is murder.'
[14] Jonathan Glover, *Causing Death and Saving Lives: The Moral Problems of Abortion, Infanticide, Suicide, Euthanasia, Capital Punishment, War and Other Life-or-Death Choices* (Penguin Books 1990) 183.
[15] Assisted Dying Bill (n 7) s 4(1).
[16] Swiss Criminal Code, 311.0; German Criminal Code; The Oregon Death with Dignity Act, Oregon Revised Statutes 127.800-995.
[17] Glover (n14) 184.
[18] ibid.
[19] Termination of Life on Request and Assisted Suicide (Review Procedures) Act 2002 (Netherlands); Annual Statutes 2016, c 3 (Canada).

both PAS and AVE and will use the term physician-assisted dying (PAD) to encompass both.[20]

Context

Moral values are often seen as being innate in humanity.[21] Natural law is one theory that accepts this view and uses it to argue that law and morality are not always concurrent, which can be regarded as proven by the fact that different societies have different laws.[22] Just because the current law does not afford individuals a certain right does not therefore mean that it should not be regarded as a human right on the basis of morality. The law is ever-changing; adapting with societal changes.[23] Whether a law is good or bad therefore often depends on how well it reflects the views of its society and its moral values.

Although there are various arguments for and against decriminalisation of PAD,[24] the purpose of this article is not to offer a summary of these arguments but rather to set out a proposal of how decriminalisation of PAD could be implemented practically and effectively in the UK. The specific conditions and safeguards proposed in this article are aimed at reducing the two main risks supposedly inherent in such decriminalisation, and so these will be briefly summarised here.

The slippery slope argument is that even if access is only granted to a small number of people to begin with, this limited scope would be difficult to justify and therefore more and more people would be granted access over time.[25] The argument is that this would

---

[20] Following Emily Jackson and John Keown, *Debating Euthanasia* (1st edn, Hart Publishing 2011). See, for differing definitions of assisted-dying, Zoe Fritz, 'The Courts should Judge Applications for Assisted Suicide, Sparing the Doctor-Patient Relationship' (*The BMJ Opinion*, 30 January 2019).

[21] Gareth Cook, 'The Moral Life of Babies: Yale Psychology Professor Paul Bloom Finds the Origins of Morality in Infants' (*The Scientific American*, 12 November 2013).

[22] Raymond Wacks, *Law: A Very Short Introduction* (OUP 2008) 81.

[23] ibid 34.

[24] For a thorough debate of the arguments surrounding the potential decriminalisation of assisted dying, see Jackson and Keown (n 20).

[25] 'Let's Not be Taken for a Ride in the Name of Compassion' (*Life Charity*, 20 May 2016) <https://lifecharity.org.uk/news-and-views/lets-not-be-taken-for-a-ride-in-the-name-of-compassion/> accessed 12 April 2021.

eventually result in the immoral taking of life without adequate restrictions and safeguards.[26] The validity of the slippery slope argument depends on presumption that 'the evidence that slippage would in fact occur however the law was framed and however the practice was regulated'.[27] The argument offers no empirical evidence for this.[28] On the contrary, the evidence from other countries which have decriminalised PAD shows that the slippery slope has not materialised; that rather that decriminalisation it has led to a steady number of procedures being carried out each year,[29] and that such laws have been safely and responsibly implemented.[30] Furthermore, any risk of a slippery slope effect can be reduced through enactment of specific restrictions and safeguards, as will be discussed throughout this article.

The fear of abuse of vulnerable patients occurring as a result of decriminalisation rests on several key arguments. The first is that some people who are pushing for decriminalisation of PAD may have ulterior motives for doing so, such as for the obtainment of body parts.[31] Although this may be true for a tiny minority, their agendas are irrelevant to the core issue of the debate. The aim of decriminalising PAD is to provide a remedy to patients who are in an

---

[26] Karen Porter and Katharine G Warburton, 'Physicians' Views on Current Legislation Around Euthanasia and Assisted Suicide: Results of Surveys Commissioned by the Royal College of Physicians' (2018) 5 Future Healthcare Journal 30.

[27] Michael Dunn and Tony Hope, *Medical Ethics: A Very Short Introduction* (2nd edn, OUP 2018) 41.

[28] ibid.

[29] Bregje D Onwuteaka-Philipsen, Arianne Brinkman-Stoppelenburg, Corine Penning and others, 'Trends in End-of-life Practices Before and After the Enactment of the Euthanasia Law in the Netherlands from 1990 to 2010: A Repeated Cross-sectional Survey' (2012) 380 The Lancet 908. Also see Helen O'Shaughnessey, 'Dying with Dignity: The Importance of Choice at End of Life' (*TEDx Talks*, 14 December 2018) <www.youtube.com/watch?v=GGUwknP0xkA&list=TLPQMDEwMzIwMjHbUT3J0rArWQ&index=11> accessed 3 May 2022: research in Oregon has shown no signs of a slippery slope effect or coercion. 80% of the public would vote for the current law again.

[30] Grace Pastine, 'Death with Dignity' (*TEDx Talks*, 19 June 2015) <www.youtube.com/watch?v=i9Q3ohzB25I&list=TLPQMDEwMzIwMjHbUT3J0rArWQ&index=12> accessed 3 May 2022.

[31] *Let's Not be Taken for a Ride* (n 25).

unbearable and irremediable amount of pain and suffering, to enable them to end their suffering in a dignified and relatively pain-free way at a time and place of their choosing.[32] The second argument is that financial considerations and resource allocation would present certain pressures for vulnerable patients, and that medically endorsed killing may become routine, and even expected of some patients, as a result.[33] The concern is that the personal autonomy of patients making the decision to undergo PAD would thereby by compromised.[34] Although this is a valid concern, research carried out in countries that have decriminalised PAD shows little or no materialisation of this risk. One example is a study carried out in Oregon, where PAD was decriminalised in 1997, which found no evidence of a disproportionate impact on patients in vulnerable groups.[35] Another study, conducted in the Netherlands (which legalised PAD in 2002)[36] has also shown no evidence of abuse.[37] Although this research does not prove that there is no risk of abuse,[38] it does show that strict safeguards and restrictions can significantly reduce it.

Why the Law Needs to Change

Some patients who have serious and debilitating conditions that cause them unbearable pain and suffering see no reason to carry on living and want to die in an autonomous and dignified way. These patients may have no potential treatment options and even the best palliative care may not alleviate their suffering. Although suicide is a viable

[32] Jackson and Keown (n 20), Series Editor's Preface: 'this may entail allowing people who cannot be helped in any other way, and who believe that death offers the only possible release from their suffering, to have their lives ended quickly and painlessly.'

[33] 'Assisted Suicide and the Hospice Alternative' (*Life Charity*, 30 November 2016) <lifecharity.org.uk/news-and-views/assisted-suicide-hospice-alternative/> accessed 12 April 2021.

[34] Imogen Goold and Jonathan Herring, *Great Debates in Medical Law & Ethics* (2nd edn, Palgrave 2018) 244.

[35] Margaret P Battin, Agnes van der Heide, Linda Ganzini and others, 'Legal Physician-assisted Dying in Oregon and the Netherlands: Evidence Concerning the Impact on Patients in 'Vulnerable' Groups' (2007) 33 Journal of Medical Ethics 591.

[36] Termination of Life on Request Act (Netherlands) (n 19).

[37] Onwuteaka-Philipsen and others (n 29) 908-915.

[38] See the conclusion reached in *Nicklinson v Ministry of Justice* [2014] UKSC 38 [88] (Lord Neuberger P).

option for some of these patients,[39] it is undeniably not the most pleasant way to die.[40] Some patients may also not have the physical or mental ability to commit suicide. The only way to end their suffering is therefore to get assistance in ending their life, which is currently prohibited in the UK.[41] As summarized by Emily Jackson in *Debating Euthanasia*, the current law 'condone[s] bringing about death by treatment withdrawal while at the same time absolutely prohibiting bringing about death in a quicker and perhaps less distressing way'.[42]

The first section of this chapter will examine the issues resulting from the current law, focusing on those impacting patients and their loved ones. The second section will look at the changing attitudes towards decriminalisation both within the UK and internationally, establishing that support has been steadily increasing amongst both the public and the medical profession. The final section examines the previous reluctance to change the law in the UK and outlines the bill currently making its way through parliament.

Issues with the Current Law

The Value of Personal Autonomy

The right to personal autonomy is regarded as an essential and basic human right in the modern world. It has been confirmed by the European Court on Human Rights as being encompassed under Article 8 of the European Convention on Human Rights.[43] It is also

[39] Suicide Act (n 3) s 1.
[40] John Bingham, 'Assisted Dying: More than 300 Terminally Ill People a Year Committing Suicide' *The Telegraph* (London, 15 October 2014) <www.telegraph.co.uk/news/uknews/assisted-dying/11163992/Assisted-dying-more-than-300-terminally-ill-people-a-year-committing-suicide.html> accessed 3 May 2022: an estimated 300 suicides a year involve people with a terminal illness.
[41] See 'Introduction', above.
[42] Jackson and Keown (n 20) 19.
[43] 'Guide on Article 8 of the European Convention on Human Rights: Right to Respect for Private and Family Life' (*Council of Europe*, 31 August 2021) <www.echr.coe.int/Documents/Guide_Art_8_ENG.pdf> accessed 5 May 2022, para 74.

firmly embedded in UK law,[44] supposedly inviolably.[45] The right was depicted as unconditional by Lord Donaldson MR in the case of *Re T*:

> [The] right of choice is not limited to decisions which others might regard as sensible. It exists notwithstanding whether the reasons for making the choice are rational, irrational, unknown or even non-existent.[46]

However, when applied to medical decisions, this right only extends to refusing treatment, not to demanding it.[47] This is because the choice over one's medical treatment is in fact a liberty rather than a right. The current law on PAD reflects this as a patient can refuse life-saving treatment but they cannot demand to receive life-ending assistance.[48] This inconsistency is problematic because it effectively limits patients' options to either committing suicide through refusing treatment such as clinically-assisted nutrition and hydration, which is self-starvation, or to continue living with their pain and suffering.

No Dignified, Autonomous Way to Die

The current law denies patients the right to die in an autonomous and dignified way. For patients who are physically or mentally incapable of ending their own life, assistance is required.[49] Without such assistance the only option left for these patients is often methods of self-starvation, as mentioned above. Tony Nicklinson was one such patient, who was left paralysed following a stroke and ended his life through self-starvation after the High Court refused to grant a declaration that it would be legal for a doctor to either kill him or assist him in killing himself.[50] Furthermore, patients who still have the physical ability to commit suicide may prefer to undergo PAD. Perhaps they are mentally incapable of carrying out the final act

[44] Human Rights Act 1998, Article 8.
[45] *F v West Berkshire HA* [1990] 2 AC 1 [72] (Lord Goff).
[46] *Re T* [1992] EWCA Civ 18 [102].
[47] Goold and Herring (n 34) 267.
[48] *Nicklinson v Ministry of Justice* (n 38) [225] (Lord Sumption) and Goold and Herring (n 34) 243 and 257.
[49] Rob Heywood and Alexandra Mullock, 'The Value of Life in English Law: Revered but not Sacred?' (2016) 36 Journal of Legal Studies 672.
[50] *Nicklinson v Ministry of Justice* (n 38).

themselves or they may prefer to die in a more dignified and relatively pain-free way, surrounded by their loved ones. The current law denies patients such autonomy and dignity in dying by not allowing them assistance. As summarised by former DPP Sir Keir Starmer:

> [w]e have arrived at a position where compassionate, amateur assistance from nearest and dearest is accepted but professional medical assistance is not, unless someone has the means and physical assistance to get to Dignitas. That to my mind is an injustice that we have trapped within our current arrangement.[51]

Premature Deaths

Another issue is that the current law leads to some patients feeling pressured to end their lives before they are ready to. Patients who wish to die in a dignified and autonomous way may fear that once their condition progresses they will not have the physical ability to end their life without assistance. They then have to decide whether to end their life prematurely or risk not having the death they desire. One example of such a patient was Richard Selly, who was unable to walk, speak, or swallow as a result of suffering from motor neurone disease. Selly decided to travel to a clinic in Switzerland in order to end his life but communicated beforehand that he was choosing to die 'earlier than [he] would prefer'.[52] The current law therefore effectively serves to cut short some patients' lives, as they fear losing the capacity to either make such a journey or commit suicide.[53] Decriminalising PAD would remove this fear as these patients would have the option of assistance available to them when they reach a stage of requiring it.

[51] *Dignity in Dying* (n 11).
[52] Becky Johnson, 'Assisted Dying could be Legalised in the UK within Four Years, Leading MP Predicts' (*Sky News*, 24 August 2020) <news.sky.com/story/assisted-dying-could-be-legalised-in-the-uk-within-four-years-leading-mp-predicts-12055523> accessed 3 May 2022.
[53] *Nicklinson v Ministry of Justice* (n 38) [96] (Lord Neuberger P).

Suicide Tourism

A further issue is that the criminalisation of PAD has resulted in patients being forced to travel to foreign countries in order to be able to die in an autonomous and dignified manner. Although some countries restrict access to PAD to residents,[54] countries such as Switzerland offer PAD to non-residents. A study into suicide tourism found that a fifth of the people who travelled to Switzerland between 2008 and 2012 to end their lives were from the UK,[55] showing that the law in the UK is not addressing the needs of its society. This is indicative of the need for change.[56] Furthermore, travelling to Dignitas or other clinics abroad costs money. Some patients simply do not have the financial resources to make the trip.[57]

Risk of Prosecution

The final issue is brought about by the criminalisation of assistance given by loved ones to patients wanting to die. Patients who accept such help, either to commit suicide or travel abroad for PAD, are risking the prosecution of their loved ones.[58] They may not be capable of taking the journey alone and therefore have to choose between the continuation of their suffering or the risk of prosecution of their loved ones. Even if they are physically capable of travelling alone, they may prefer to be joined by loved ones for company and support.

---

[54] Oregon Death with Dignity Act (n 16) s 2.01(1).
[55] Saskia Gauthier, Julian Mausbach, Thomas Reisch and others, 'Suicide Tourism: A Pilot Study on the Swiss Phenomenon' (2015) 41 Journal of Medical Ethics 611.
[56] Haroon Siddique, 'One in Five Visitors to Swiss Assisted-Dying Clinics from Britain' *The Guardian* (London, 20 August 2014) <www.theguardian.com/society/2014/aug/20/one-in-five-visitors-swiss-suicide-clinics-britain-uk-germany> accessed 3 May 2022.
[57] Access to Dignitas services requires payment of a one-off joining fee of 200 CHF as well as an annual membership fee of 80 CHF: 'Objectives and purpose of DIGNITAS' (*Dignitas*) <www.dignitas.ch/index.php?option=com_content&view=article&id=22&Itemid=5&lang=en> accessed 3 May 2022.
[58] Raymond Tallis, 'Assisted Dying: Why the RCP should be Neutral' (*Royal College of Physicians*, 14 January 2019) <www.rcplondon.ac.uk/news/assisted-dying-why-rcp-should-be-neutral> accessed 3 May 2022.

Although assisting a suicide is illegal, it is rare for someone to be prosecuted successfully for it if they had good intentions. Since 2009, 138 assisted suicide cases have been referred to the CPS, 19 of which were prosecuted.[59] None resulted in convictions, thanks to the common-sense approach adopted by the courts.[60] DPP policy states that it is unlikely that someone acting wholly out of compassion would be prosecuted for assisting a person who made a voluntary and informed decision to end their life.[61] Nevertheless, this is not a guarantee and does not eliminate the risk of prosecution completely.

Changing Attitudes

Views on assisted dying are shifting across the globe and the number of countries that have decriminalised PAD is increasing.[62] In the UK in recent years, there have been major changes in the views of the public as well as medical practitioners.[63] This evidences a societal shift calling for a change in the law.[64]

The Public

A 2016 British Social Attitudes survey found that 78% of the British public believed that PAD should be decriminalised where it is carried out by a physician and where the patient is suffering from a painful and incurable terminal illness.[65] The survey also found strong support (51%) for decriminalising PAD in cases where the patient is not

[59] Alexa Payet, 'The Tragedy of Assisted Suicide' (*Bolt Burdon Solicitors*) <www.boltburdon.co.uk/blogs/assisted-suicide-will-dispute/> accessed 3 May 2022.
[60] ibid.
[61] DPP, 'Suicide: Policy for Prosecutors in Respect of Cases of Encouraging or Assisting Suicide' (*CPS*, October 2014) <www.cps.gov.uk/legal-guidance/suicide-policy-prosecutors-respect-cases-encouraging-or-assisting-suicide> accessed 5 May 2022.
[62] For example, Organic Law for the Regulation of Euthanasia (Spain) and BVerfG, *Judgment of the Second Senate of 26 February 2020*, 2 BvR 2347/15 (Germany).
[63] See Tallis (n 58).
[64] There is a significant link between claims to certain rights and the establishment of new laws: Wacks (n 22) 84.
[65] British Social Attitudes, 'Moral Issues: Sex, Gender Identity and Euthanasia' (34th edn, NatCen Social Research 2016) 25.

terminally ill.[66] This suggests that any reform of the law would need to enable access to PAD for non-terminal as well as terminal patients in order to satisfy public demand. This will be discussed further in chapter two.

Support for decriminalisation is also increasing amongst members of parliament. This is evidenced by calls for the government to hold an inquiry into the current law, which argue that the UK has fallen behind the rest of the world.[67] Norman Lamb MP has stated that, 'if someone acts out of absolute compassion, they are still left with their home being declared a crime scene and with a police investigation... That is surely an intolerable position.'[68] Although previous bills have been unsuccessful, Andrew Mitchell MP has stated his belief that, with certain restrictions in place, parliament may support the decriminalisation of PAD in the near future.[69]

Practitioners

Support for decriminalising PAD is also growing amongst members of the medical profession. The traditional position has been to oppose any legal change but in recent years this has shifted towards a more neutral stance. The Royal College of Nursing adopted a neutral stance on PAD in cases of terminal illness in 2009.[70] The Royal College of Physicians (RCP) adopted a neutral stance on assisted dying in 2019, following a survey which revealed that 31.6% of their members

[66] ibid.
[67] Helena Wilkinson, 'Assisted Dying Inquiry Essential, Leading Brain Surgeon Says' (*BBC News*, 1 April 2021) <www.bbc.co.uk/news/uk-56597771> accessed 3 May 2022.
[68] *Dignity in Dying* (n 11).
[69] Johnson (n 52).
[70] 'RCN Position Statement on Assisted Dying' (*Royal College of Nursing*, 6 November 2014) <www.rcn.org.uk/about-us/our-influencing-work/policy-briefings/pol-2314> accessed 3 May 2022.

thought that the RCP should support a change in the law,[71] and 40.5% personally supported such a change.[72]

By contrast, the British Medical Association (BMA) reaffirmed its stance opposing decriminalisation of PAD in 2016.[73] However, a survey carried out in February 2020 revealed that 40% of its members thought that the BMA should support a change in the law and 50% personally believed that there should be such a change,[74] suggesting that the BMA may soon assume a neutral stance.[75] Nevertheless, the Royal College of General Practitioners reaffirmed their opposition in 2020, despite a member consultation revealing that 40% of respondents supported a change in the law.[76]

Previous Reluctancy to Change the Law

Despite increasing support for decriminalisation of PAD in the UK, the law remains unchanged. However, shifting societal attitudes have to some extent been reflected through decisions made by the courts.[77] In the case of *Airedale NHS Trust v Bland*,[78] it was held that it was not

---

[71] 'The RCP Clarifies its Position on Assisted Dying' (*Royal College of Physicians*, 26 March 2020) <www.rcplondon.ac.uk/news/rcp-clarifies-its-position-assisted-dying> accessed 3 May 2022.

[72] 'No Majority View on Assisted Dying Moves RCP Position to Neutral' (*Royal College of Physicians*, 21 March 2019) <www.rcplondon.ac.uk/news/no-majority-view-assisted-dying-moves-rcp-position-neutral> accessed 3 May 2022.

[73] 'The BMA's Position on Physician-assisted Dying' (*British Medical Association*, 28 September 2020) <www.bma.org.uk/advice-and-support/ethics/end-of-life/the-bmas-position-on-physician-assisted-dying> accessed 3 May 2022.

[74] *BMA Survey on Physician-Assisted Dying: Research Report* (Kantar, Public Division 2020) 3.

[75] However, the BMA consultation only had a 19.35% response, so it cannot be assumed that the BMA's position will change: - 'Physician-assisted Dying Survey' (*British Medical Association*, 20 January 2021) <www.bma.org.uk/advice-and-support/ethics/end-of-life/physician-assisted-dying-survey> accessed 3 May 2022.

[76] 'Assisted Dying: RCGP's 2020 Decision' (*Royal College of General Practitioners*) <www.rcgp.org.uk/policy/rcgp-policy-areas/assisted-dying.aspx> accessed 3 May 2022. The consultation only had a 13.47% response, so their decision is arguably justified.

[77] John Coggon, 'The Wonder of Euthanasia: A Debate that's being Done to Death' (2013) 33 Oxford Journal of Legal Studies 403.

[78] [1993] AC 789.

in the best interests of a person in a permanent vegetative state to be kept alive and that physicians could therefore withdraw life-sustaining treatment. In the case of *An NHS Trust and others v Y and another*,[79] the Supreme Court ruled that a court's permission is not required for the withdrawal of treatment in cases where the patient is in a permanent vegetative state and if the physicians and family of the patient agree that withdrawal is in the patient's best interests. However, the courts have been reluctant to create legal precedent for medical assistance to die that goes beyond withdrawal of life-sustaining treatment. In *Newby v. Secretary of State for Justice*,[80] it was considered that creating such precedent would go far beyond the courts' powers and duties, which consist of analysing evidence and interpreting pre-established law. The courts seem to be 'anxious to avoid allegations of usurping parliamentary sovereignty.'[81] Any further movement towards the decriminalisation of PAD is therefore likely to require action by parliament.

Factors to be Considered

This chapter will look at the factors that would need to be considered for the decriminalisation of PAD to be effective at preventing a slippery slope effect. First, it will determine the eligibility requirements that would need to be enforced, including life expectancy (i.e. whether a patient is terminal or not), mental as well as physical conditions, and age restraints. It will then establish the threshold conditions that would need to be implemented, including the treatability of patients' conditions, their level of pain and suffering, and their capacity to consent.

[79] [2018] UKSC 46.
[80] [2019] EWHC 3118 (Admin). The same justification was given in *Nicklinson v Ministry of Justice* (n 38) and *Conway v Secretary of State for Justice* [2018] EWCA Civ 1431.
[81] Heywood and Mullock (n 49) 675.

Who Should be Eligible?

Life Expectancy

Several countries that have decriminalised PAD have limited access to PAD to patients with a terminal illness.[82] This is the position in the Assisted Dying Bill.[83] One argument for restricting access to PAD to terminal patients is the potential for the discovery of new treatments which could alleviate or even end the suffering of non-terminal patients. However, new treatments may also be discovered for terminal conditions which render them non-terminal. This point is made far less frequently, somewhat undermining the argument.

Another argument, one expressed by ethicist Theo Boer, is that non-terminal patients have a greater opportunity to find new balance and meaning in life due to their longer life expectancy, and that they should be given this opportunity.[84] The issue with these arguments is that one of the main reasons for decriminalisation is to grant people personal autonomy over their death, and so the decision of whether to die, wait for a potential new treatment or adopt a new balance in life should rest with the patient.

A further argument supporting this restrictive approach is that terminal patients are already dying and cannot therefore *choose* to die in any meaningful way. Rather, they are choosing how and when to die; trying to get back a little sense of control and autonomy.[85] The obvious criticism of this argument is that in some sense everyone is dying and that terminally ill people are just closer to death than most others. *A fortiori*, the question is how long a patient has left to live

---

[82] Oregon Death with Dignity Act (n 16) s 1.01(12).
[83] Assisted Dying Bill (n 7) s 2.
[84] Christopher de Bellaigue, 'Death on Demand: Has Euthanasia Gone Too Far?' *The Guardian* (London, 18 January 2019) <www.theguardian.com/news/2019/jan/18/death-on-demand-has-euthanasia-gone-too-far-netherlands-assisted-dying> accessed 3 May 2022.
[85] Lecretia Seales and Matt Vickers, 'The Choices We have about How We Die' (*TEDx Talks*, 9 November 2015) <www.youtube.com/watch?v=1c7DMcgMw0o&list=TLPQMDEwMzIwMjHbUT3J0rArWQ&index=14> accessed 3 May 2022. See around 10:50.

rather than whether they are already dying.[86] The dilemma then becomes whether a long life holds more value than a short one. However, arguably misses the crux of the matter. The central consideration should be a patient's quality of life, and therefore the level of suffering they are experiencing, as opposed to length of life. The level of a patient's suffering is not dependent on whether their condition is terminal or not.[87]

Any criterion of terminal illness for access to PAD would require a quantification of the expected lifespan of terminal patients. Other legislatures, such as those of New Zealand and Oregon, have defined a terminally ill person as a person who is reasonably expected to die within six months.[88] Canada has adopted the approach of requiring that a patient's natural death has become 'reasonably foreseeable'.[89] One argument against implementing a specific timeframe is that physicians' estimates of a patient's life expectancy are inherently uncertain. Patients often live for longer or shorter periods than predicted,[90] and so adopting a terminal illness criterion with a specific timeframe is not reliable.

Supporters of access to PAD for non-terminal as well as terminal patients argue that without access to PAD, non-terminal patients face a much longer time spent in pain and suffering.[91] The

---

[86] Glover (n 14) 54.

[87] ibid 55.

[88] See End of Life Choice Bill 2017 (New Zealand), s 4(1)(c); Oregon Death with Dignity Act (n 16) s 1.01(12). This is also stipulated in the definition of 'terminal illness' in the Assisted Dying Bill (n 7) s 2(1)(b).

[89] Annual Statutes 2016, c 3 (Canada) (n 19) s 241.2(2)(d).

[90] 'Don't Make Oregon's Mistake: Ten Reasons why the UK should not Follow Suit and Legalise Assisted Suicide' (*Care Not Killing*) <www.carenotkilling.org.uk/public/pdf/falconer-bill---oregon.pdf> accessed 3 May 2022;
'Oregon Assisted Suicides Jump 28%' (*Care Not Killing*, 1 March 2021) <www.carenotkilling.org.uk/articles/oregon-assisted-suicides-jump-28/> accessed 3 May 2022; and
Rob George, 'We must not Deprive Dying People of the Most Important Protection' (2014) 349 British Medical Journal.

[91] See Goold and Herring (n 34) 245.
Also see the quote by Dutch MP Pia Dijkstra in Harriet Sherwood, 'A Woman's Final Facebook Message Before Euthanasia: 'I'm Ready for my Trip now...'' *The Guardian* (London, 17 March 2018)

case of Tony Nicklinson is again illustrative. Nicklinson suffered a stroke which left him with locked-in syndrome. He was not terminally ill.[92] Denying patients like him access to PAD seems unjust and arguably unreasonable. Although public support for access to PAD for the terminally ill is stronger (78%)[93] than for those with non-terminal conditions (51%),[94] the fact that there is still majority support for the less restrictive option is significant.

With these considerations in mind, the best approach to decriminalisation in the UK is arguably to follow the Netherlands and Belgium, where eligibility for PAD is based on the patient's condition being 'irremediable and incurable', rather than terminal.[95] Although Canadian law does not require a patient to be terminal it does require them to be in 'an advanced state of decline in capability'.[96] This condition is problematic. The loss of mental capability could cause a patient's capacity to come into question, which is a necessary threshold condition that will be discussed in later in this chapter. Furthermore, the loss of physical capability could compromise a patient's ability to give proper consent if they become unable to write or speak. This safeguard will be discussed in chapter three.

The 'irremediable and incurable' test is a reasonable one for the UK to adopt.[97] This would ensure both that patients have exhausted other options before resorting to PAD and that patients who are experiencing irremediable suffering would have access to PAD regardless of their estimated life expectancy.

Physical or Psychological Suffering

The main argument for decriminalising PAD for psychological as well as physical conditions is that it should not matter whether a patient's suffering is physical or psychological. Rather, only the extent of the

<www.theguardian.com/society/2018/mar/17/assisted-dying-euthanasia-netherlands> accessed 3 May 2022.

[92] *Nicklinson v Ministry of Justice* (n 38).

[93] British Social Attitudes (n 65) 25.

[94] ibid.

[95] See The Belgian Act on Euthanasia of May, 28th 2002, s 3(1).

[96] Annual Statutes 2016, c 3 (Canada) (n 19) s 241.2(2).

[97] Heywood and Mullock (n 49) 679.

suffering should matter. Canada and Belgium have both adopted this approach and decriminalised PAD in cases of psychological as well as physical suffering.[98]

The main argument against decriminalising PAD for psychological conditions and only doing so for physical conditions, as has been done in New Zealand,[99] is focused on the nature of psychological conditions. Many psychological conditions, such as depression, are treatable, in which case the patient would not meet the requirement of having an irremediable condition. Furthermore, if a psychological condition has progressed past a certain point then the patient will no longer meet the mental capacity requirement to consent to PAD. Even if these conditions are met, some psychologically ill patients seeking PAD may not be in a state of unbearable pain and suffering. Therefore, even if access to PAD were granted to patients with psychological as well as physical conditions, many patients with psychological conditions would not meet the other necessary criteria and conditions.

Another argument against decriminalising PAD for psychological conditions is that not enough research has been done into psychological illnesses and potential treatments. Some of these illnesses may be treatable, or even curable, but we have not yet made these discoveries. This argument rests on the basis that it would not be in a patient's best interests to undergo PAD if there is the possibility of new treatments. Although this argument could also be applied to physical conditions, it does not have the same force. Most medical research concerns physical conditions, meaning our understanding of psychological illness is less advanced.[100] However, critics of this position argue that the autonomy to decide whether to wait for a potential treatment or to choose to die should be given to the patient.

---

[98] Annual Statutes 2016, c 3 (Canada) (n 19) s 241.2(2)(c) and Belgian Act on Euthanasia (n 95) s 3(1).

[99] End of Life Choice Bill (New Zealand) (n 88) s 4(1)(d).

[100] Sophie Dix, 'Filling the Gap in Mental Health Research: Ways to Tackle the Lack of Funding in Mental Health Research' *The Pharmaceutical Journal* (London, 12 November 2015) <https://pharmaceutical-journal.com/article/opinion/filling-the-gap-in-mental-health-research> accessed 6 May 2022.

Although there are additional considerations around the decriminalisation of PAD for psychological conditions, these are not enough to justify prohibiting patients with psychological pain and suffering from accessing PAD. If patients meet the other conditions proposed in this article, specifically that their condition is incurable and irremediable, and that patients have the necessary capacity to consent, then they should be granted the same access as patients with physical conditions.

Age Restriction

Most countries which have decriminalised PAD have restricted access to adults.[101] One exception is the Netherlands, where PAD is accessible to anyone over the age of twelve, taking into consideration the maturity of the child and parental consent.[102] Since age does not affect the level of pain and suffering experienced by a patient, it seems reasonable to extend access to PAD to children, as long as they meet the other eligibility criteria and threshold conditions.

However, there remain valid arguments for restricting access to PAD to adults. The first is that children generally have a longer life ahead of them. This argument again presumes that a longer life has more value than a shorter one. However, as discussed above, this is not necessarily always the case. The key factor should be the level of suffering rather than expected lifespan. The second argument is that children under a certain age often do not have enough life experience to properly process the complexities of a decision to access PAD and give well-considered and informed consent. It would therefore be reasonable to restrict access to PAD to adults over the age of eighteen,[103] as the Assisted Dying Bill proposes.[104]

[101] Annual Statutes 2016, c 3 (Canada) (n 19) s 241.2(1)(b); End of Life Choice Bill (New Zealand) (n 88) s 4(1)(a); and Belgian Act on Euthanasia (n 95) s 3(1).
[102] Termination of Life on Request Act (Netherlands) (n 19) art 2.
[103] This is the age of adulthood in the UK, per the Family Law Reform Act 1969, s 1(1).
[104] Assisted Dying Bill (n 7) s 1(2)(c)(i).

What Should the Threshold Conditions be?

Pain and Suffering

First, we must consider the level of pain and suffering being experienced. Although different countries have used slightly varied wording, they have all adopted similar thresholds. In Canada, the level of pain must be 'intolerable',[105] whilst in Belgium, New Zealand and the Netherlands the wording used is 'unbearable'.[106] These words have near-identical meaning. This article proposes the adoption of a threshold condition of 'unbearable pain and suffering'. Pain and suffering are both subjective concepts because pain thresholds vary among individuals. Fulfilment of this requirement should therefore be analysed on an individual basis dependent on the patient's own assessment.[107] The Assisted Dying Bill does not include this condition.

Another element of pain and suffering that must be considered is its continuity. The law governing PAD in Belgium requires that the patient's suffering must be 'constant'.[108] This condition is problematic as it means that a patient may not qualify for PAD if their pain and suffering is lasting but intermittent. Although some patients are in a constant state of unbearable pain and suffering, others suffer intermittently. This does not necessarily make their suffering less irremediable or unbearable. To reiterate, pain and suffering are subjective and therefore the decision over what is unbearable to individual patients should rest with them. The UK should not adopt a requirement that a patient's pain is constant.

Capacity to Consent

Any legislation decriminalising PAD in the UK would need to stipulate that a patient has the necessary mental capacity to consent in

[105] Annual Statutes 2016, c 3 (Canada) (n 19) s 241.2(2)(c).
[106] Belgian Act on Euthanasia (n 95) s 3(1); End of Life Choice Bill (New Zealand) (n 88) s 4(1)(e); and Termination of Life on Request Act (Netherlands) (n 19) art 2(1)(b).
[107] *Purdy v DPP* [2009] UKHL 45 [66] (Baroness Hale).
[108] Belgian Act on Euthanasia (n 95) s 3(1).

order to access PAD. English law encompasses the presumption that a patient has such capacity unless proven otherwise.[109] Therefore, in order to protect vulnerable patients and reduce the risk of a slippery slope effect, legislation decriminalising PAD should stipulate that physicians may request a psychological assessment of the patient if they have any doubt as to the patient's capacity. This safeguard will be discussed further in chapter three.

Capacity to consent is a necessary condition of decriminalising PAD. However, some patients may wish to use their current capacity to make an advance decision to receive PAD which is not invalidated by the future loss of capacity. The Mental Health Act 2005 states that if a patient no longer has the capacity to make a decision then whoever is making the decision on their behalf has to take into account their past feelings and wishes, particularly those which were expressed in writing whilst they still had capacity.[110] One way for a patient to express such wishes is to make an advance directive. This option would be especially significant for patients with progressive neuro-degenerative conditions, such as Alzheimer's disease or dementia, who want PAD in the future but fear that they will lose the necessary mental capacity to give proper consent. In the Netherlands, advance consent can be given for undergoing PAD.[111] The patient has to clearly state under which circumstances they wish to be granted euthanasia and what they personally deem 'unbearable suffering'.[112]

However, due to UK law restricting the right of decision-making to refusal of treatment, advance directives would only be an option for patients who wish to refuse PAD and not those who wish to undergo it. Furthermore, the Assisted Dying Bill proposes that the

---

[109] Mental Capacity Act 2005, s 1(2). Also see *Re MB* [1997] EWCA Civ 3093 [30] (Butler-Sloss LJ).

[110] Mental Capacity Act (n 109) s 4(6)(a).

[111] 'Euthanasia: Dutch Court Expands Law on Dementia Cases' (*BBC News*, 21 April 2020) <www.bbc.co.uk/news/world-europe-52367644> accessed 3 May 2022.

[112] Eva C A Asscher and Suzanne Van de Vathorst, 'Supreme Court Rules on the First Prosecution of a Dutch Doctor Since the Euthanasia Act' (*Journal of Medical Ethics Blog*, 28 April 2020). <https://blogs.bmj.com/medical-ethics/2020/04/28/supreme-court-rules-on-the-first-prosecution-of-a-dutch-doctor-since-the-euthanasia-act-2/> accessed 8 May 2022.

current, informed consent of the patient should be required for PAD,[113] thereby disallowing the option of advance consent. Although this removes the option of requesting PAD in advance for people who fear losing capacity, it does provide an important protection for those who do not want PAD as they could make an advance directive stating their wish to refuse PAD if they lose capacity in the future.

Safeguards to be Enforced

Any future legislation would have to implement certain safeguards to protect patients who may be vulnerable to abuse or external pressures, as well as physicians who either want to protect themselves against claims of abuse or clinical negligence or who do not want to participate in PAD. This chapter will first look at the safeguards needed to protect patients, before discussing those necessary to protect physicians.

Safeguards for Patients

In order to protect vulnerable patients from the risk of being killed against their will, certain safeguards would be required. These should be designed to ensure that any decision to undergo PAD is truly autonomous, without external pressures or influences. They should also ensure that the patient is reaching this decision after receiving all the necessary information and giving it proper consideration, and that they have the mental capacity to give valid consent.

Patient's Initial Request

One of the main fears of decriminalising PAD is that it would lead to a culture where choosing to die is normalised and even expected of certain individuals because they are no longer 'useful' to society or are a burden on their family or on the healthcare system.[114] Such pressure could arise externally from others or from society as a whole. It could equally arise internally, where an individual wants to end their life because they feel that they have become a burden and as a result they

[113] Assisted Dying Bill (n 7) s 1(2)(c)(ii).
[114] *Nicklinson v Ministry of Justice* (n 38) [89] (Lord Sumption) and Goold and Herring (n 34) 249.

feel an obligation to end their life.[115] The question that then arises is one of true autonomy.[116] Even if a patient was requesting PAD on the basis of feeling like a burden, this does not mean that we should not respect their autonomy to make that decision. If a patient sees that they are putting great strain on their family or loved ones and decides that they would rather die than continue in this situation, this does not render their decision invalid or mean that it is not autonomous and well-considered.[117] It is simply one of the factors they are taking into account. In any case, in order for patients to access PAD, they would need to be in irremediable, unbearable pain and suffering. The additional restrictions and safeguards being proposed would therefore limit the overall effect that such pressure could potentially have on vulnerable patients as they would also need to meet the other eligibility criteria and threshold conditions.

Nevertheless, any legislation decriminalising PAD should aim to minimize the risk of such pressures on patients. Although the law cannot prevent family members or loved ones from putting certain pressures on a patient, it can prevent physicians from doing so by specifying that they cannot suggest PAD or start the conversation with patients. Although the Assisted Dying Bill does not include such a safeguard, it has been adopted by several other countries and should be included in any future UK legislation.[118] This would mean that the request for PAD would have to be initiated by the patient, ensuring that the wish to end their life stems from a true desire to die rather than external pressures.

In order to ensure that this safeguard would be effective, it would also be necessary for the law to outline how such an initial request should be made. In Belgium, a request for PAD must be made in writing,[119] and the new bill currently making its way through the Spanish parliament contains the same requirement.[120] Requiring

[115] Kate Greasley, 'R (Purdy) v DPP and the Case for Wilful Blindness' (2010) 30 OJLS 301.
[116] Goold and Herring (n 34) 244.
[117] Glover (n 14) 187.
[118] For example, End of Life Choice Bill (New Zealand) (n 88) s 7(1).
[119] Belgian Act on Euthanasia (n 95) s 3(4).
[120] Minder (n 62).

written requests helps ensure that proper, informed consent[121] is being given by the patient.

That proper, independently formed consent has been given is of overarching importance in any case. This should be stipulated separately in any legislation decriminalising PAD, as is the case in other jurisdictions.[122]

One way of ensuring that a patient is giving well-considered, autonomous consent is requiring reiteration of the request.[123] Other jurisdictions have adopted this safeguard, stipulating that the reiterated request be made directly prior to fulfilment.[124] Requiring a reiterated request would provide further protection for both vulnerable patients as well as patients who simply change their mind about wanting to die, whilst also ensuring that patients are fully aware of their option to withdraw consent at any point.[125] The Assisted Dying Bill includes this safeguard as it stipulates that physicians would be required to check that the patient does not revoke their consent directly prior to fulfilment.[126]

Cooling-off Period

The period between the initial request for PAD and the request being reiterated and fulfilled is called a cooling-off period. This period ensures that when carrying out PAD, physicians would be abiding by the patient's true wishes rather than an impulsive decision stemming from a temporary lack of will to live.[127] In order to be effective, the cooling-off period would need to give patients enough time to

---

[121] 'Decision Making and Consent: The Dialogue Leading to a Decision' (*General Medical Council*, 9 November 2020) <www.gmc-uk.org/ethical-guidance/ethical-guidance-for-doctors/decision-making-and-consent/the-dialogue-leading-to-a-decision#paragraph-10> accessed 3 May 2022.

[122] Oregon Death with Dignity Act (n 16) s 1.01(7); Belgian Act on Euthanasia (n 95) s 3(1) and (2); Termination of Life on Request Act (Netherlands) (n 19), arts 2(1)(a) and (c).

[123] As in Belgium: Belgian Act on Euthanasia (n 95) s 3(1).

[124] Oregon Death with Dignity Act (n 16) s 3.01(i); Termination of Life on Request Act (Netherlands) (n 19), arts 2(1)(a) and (c); Belgian Act on Euthanasia (n 95) s 3(1).

[125] Glover (n 14) 185.

[126] Assisted Dying Bill (n 7) s 4(2)(c).

[127] Goold and Herring (n 34) 248 and Glover (n 14) 185.

properly consider the benefits and consequences of their decision. At the same time, it cannot be so long as to significantly increase the risk of the patient's condition deteriorating, particularly if there is a risk that the patient loses the capacity to consent. There is no general standard for how long this period should be. Different jurisdictions have set different time limits. For instance, Oregon has a fifteen-day waiting period;[128] Canada's is ten days.[129] Belgium only has a time limit (one month) in place for non-terminal patients.[130] The approach of implementing a separate time limit for non-terminal patients is problematic if it is accepted that access to PAD should be based on the level and extent of a patient's pain and suffering rather than their life expectancy. The Assisted Dying Bill (which would restrict access the terminally ill) stipulates a cooling-off period of fourteen days,[131] which is consistent with other UK law.[132]

Psychological Assessment

The legal presumption that a patient has mental capacity[133] creates a potential safeguarding issue in respect of PAD. There is a risk that patients may not in fact have the requisite capacity, meaning they undergo PAD without having properly consented. One possible safeguard would be to require a patient to undergo a psychological assessment if their physician has any reason to doubt their mental capacity. This safeguard has been adopted in New Zealand and Oregon,[134] and is contained in the Assisted Dying Bill.[135] Under the bill, if the results of the assessment show that the patient does not in fact have the necessary mental capacity to consent to PAD, then they would be denied access.

[128] Oregon Death with Dignity Act (n 16) s 3.08(1).
[129] Annual Statutes 2016, c 3 (Canada) (n 19) s 241.2(3)(g).
[130] Belgian Act on Euthanasia (n 95) s 3(3)(2).
[131] Assisted Dying Bill (n 7) s 4(2)(d).
[132] For example, The Consumer Contracts (Information, Cancellation and Additional Charges) Regulations 2013 para 30(2).
[133] Mental Capacity Act (n 109) s 1(2).
[134] End of Life Choice Bill (New Zealand) (n 88) s 12; Oregon Death with Dignity Act (n 16) s 3.03.
[135] Assisted Dying Bill (n 7) s 13(5).

Although this safeguard is useful, it is imperfect in that that it leaves the decision whether or not to request a psychological assessment up to the physicians, who may not realise that a patient's capacity is declining. The possibility remains that access to PAD is granted to patients who do not have the capacity to properly consent. However, any stricter safeguard would be problematic in that it would conflict with the current presumption under English law that a person has capacity.

Safeguards for Physicians

In order to protect physicians from accusations of abuse or negligence and to protect physicians from being forced to participate in PAD against their will, certain safeguards would have to be implemented.

Two Physicians' Agreement

Any future decriminalisation of PAD in the UK should require that two physicians have to agree on the patient's condition, their prognosis, and reach the conclusion that they meet the necessary eligibility criteria and threshold conditions for access to PAD. This safeguard is well established in UK medical law, for example under the Abortion Act,[136] for protecting patients from the effects of abuse or clinical negligence as well as protecting physicians from being accused of such acts.

Another presumed advantage of this safeguard is that it would reduce the risk of a slippery slope effect by ensuring that the eligibility requirements and threshold conditions for accessing PAD would be complied with. However, critics argue that this presumption is flawed. They cite UK abortion law, where the two-doctor safeguard is thought to have failed to prevent on-demand abortion because physicians pre-sign the required forms without actually seeing or assessing the patient.[137] However, the risk of this happening with PAD could arguably be diminished by stipulating that physicians have to be independent from one another and that they both have to visit and

[136] Abortion Act 1967, s 1(1).

[137] Sally Sheldon, Gayle Davis, Jane O'Neill, and others, 'The Abortion Act (1967): A Biography' (2019) 39 Journal of Legal Studies 25.

assess the patient in-person in order to satisfy the condition. Another requirement that could be implemented to strengthen this safeguard is that both of the physicians' findings must be confirmed in the form of a written report, so as to ensure proper compliance. This requirement has also been implemented in several foreign PAD laws.[138]

In order to effectively implement this safeguard and ensure that the other conditions are fulfilled in practice, legislation would need to stipulate that:

(1) the physicians are independent from one another;
(2) that both the attending and consulting physician assess the patient separately and reach their own conclusion as to the patient's condition; and
(3) that both physicians have to produce a written report.

The first two of these conditions are stipulated in the Assisted Dying Bill,[139] and can be found in numerous foreign PAD decriminalisation laws.[140] However, the third condition is not included in the bill but should be included in any future UK legislation to ensure proper safeguarding of physicians.

Conscientious Objection

As previously outlined, the British medical profession has traditionally opposed any decriminalisation of PAD. Although this position has become more neutral in recent years, hesitancy remains amongst physicians. This is evident from the BMA's reaffirmation of its stance against decriminalising PAD in 2016.[141] The aim of prolonging life has been inherent to the medical profession since the adoption of the Hippocratic Oath. This is likely one reason why some

[138] For example, Annual Statutes 2016, c 3 (Canada) (n 19) s 241.2(3)(f); Oregon Death with Dignity Act (n 16) s 3.02; and Termination of Life on Request Act (Netherlands) (n 19) art 2(1)(e).
[139] Assisted Dying Bill (n 7) s 3(3).
[140] Annual Statutes 2016, c 3 (Canada) (n 19) s 241.2(3)(f); Termination of Life on Request Act (Netherlands) (n 19) art 2(1)(e); End of Life Choice Bill (New Zealand) (n 88) s 10 & 11; Belgian Act on Euthanasia (n 95) s 3(2)(3).
[141] The BMA's Position on Physician-assisted Dying (n 73).

physicians are still hesitant about PAD. Like abortion, PAD necessitates doctors actively ending or helping to end lives.[142] However, it has been argued that the primary aim of physicians should be to relieve the pain and suffering of their patients and therefore focus more on the quality of life rather than strictly the preservation of it.[143] This argument was well illustrated by Steven Pleiter, director of Levenseindekliniek, a Dutch end-of-life clinic:

> if the situation is unbearable and there is no prospect of improvement, and euthanasia is an option, it would be almost unethical [of a doctor] not to help that person.[144]

If a person is experiencing an unbearable amount of pain and suffering and requires assistance in order to end their suffering, the most ethical approach would be to help them to do so. However, this is difficult in the context of PAD because it would require physicians to supply or administer patients with life-ending drugs.

Some physicians may have very good ethical, religious, or other reasons for not wanting to partake in PAD. A survey carried out by the British Medical Association in 2020 found that only 36% of physicians would willingly participate in the prescription of life-ending drugs, and only 26% would be willing to actually administer such drugs to patients.[145] So, although it is necessary for physicians to be the ones carrying out PAD in order to reduce the risk of abuse, we cannot expect all physicians to be willing to do so.[146]

---

[142] 'Treatment and Care Towards the End of Life: Good Practice in Decision Making' (*General Medical Council*, 20 May 2010) <www.gmc-uk.org/-/media/documents/treatment-and-care-towards-the-end-of-life---english-1015_pdf-48902105.pdf?la=en&hash=41EF651C76FDBEC141FB674C08261661BDEFD004> accessed 3 May 2022, para 10.

[143] John Chisholm, chair of the BMA medical ethics committee, quoted in in Sarah Johnson, 'New Rules Spell out when Doctors can let Patients with Brain Damage Die' *The Guardian* (London, 12 December 2018) <www.theguardian.com/society/2018/dec/12/dying-guidance-patients-persistent-vegetative-state> accessed 3 May 2022.

[144] De Bellaigue (n 84). Also see the quotation by Professor Raymond Tallis: 'we have a moral and clinical duty to respect the wishes of our patients', Tallis (n 58).

[145] *BMA Survey on Physician-Assisted Dying* (n 74) 3.

[146] Porter and Warburton (n 26) 30-34.

Any future legislation would therefore need to take physicians' varied views into account and incorporate a right to conscientious objection for doctors.[147] This should be similar to that given to physicians in cases of abortion.[148] The Assisted Dying Bill incorporates this right.[149]

Although physicians would need to be afforded the right to conscientiously object to carrying out PAD procedures, this could create issues for patients, as some may thereby be prevented from accessing PAD. Another necessary safeguard would therefore be to place a positive obligation on conscientiously-objecting physicians to refer patients to another, willing physician.

Practical Application

Having considered the necessary conditions and safeguards for decriminalising PAD, this chapter will outline how such legislation could be applied practically and efficiently. Crucially, who should be given the decision-making power over patient requests for PAD? This chapter will discuss the three different options: physicians, the courts, or an arms-length authority.

Whichever approach is adopted, annual reporting and review requirements are essential to safeguarding against the risks of decriminalisation. Publication of key information on PAD ensures transparency and accountability for decision-makers. Report and review requirements have been adopted by several other countries and have been instrumental in safeguarding.[150] Such requirements also play a key role in maintaining public confidence in PAD.

---

[147] Some doctors support decriminalisation of PAD but would not participate personally in it: Peter T Hetzler and others, 'A Report of Physicians' Beliefs about Physician-assisted Suicide: A National Study' (2019) 92 Yale Journal of Biology & Medicine 575.

[148] Abortion Act (n 136) s 4.

[149] Assisted Dying Bill (n 7) s 5.

[150] Oregon Death with Dignity Act (n 16) ss 3.11(1)(a), (b) and 3.11(3); End of Life Choice Bill (New Zealand) (n 88) s 17.

The Courts

The first option would be for the decision-making power over individual applications for PAD to be entrusted to the courts. This model entails patients making applications for PAD to the courts, which would evaluate and decide each case.[151] One key disadvantage of this approach court proceedings tend to be long and drawn out. Some terminally ill patients may die before their application has been considered. One way in which this issue might be addressed would be to require that PAD applications are fast-tracked, prioritising them over other less time-sensitive matters.

Another issue is that few judges are experts in PAD. They would therefore have to rely heavily on the declarations made by the two physicians. Handing first instance responsibility for PAD applications to the courts would result in a more costly and time-consuming process than allowing physicians to make the decision themselves. It would also consume court time and resources for which there is already strong demand. On balance, the courts are not the best place for considering PAD applications in the first instance.

Physicians

The second option would be to leave the decision-making to physicians. This would result in a process whereby a patient seeking PAD would make the initial request, specifying whether they would like to receive a prescription or whether they want a physician to carry out the final act. The attending physician would then evaluate whether or not the patient meets the legislative requirements before consulting another, independent physician who would do the same. If both the physicians agree that the patient meets the requirements, then arrangements would be made for the patient's for PAD to be fulfilled. This would either be through prescription of the life-ending drug to the patient, which could be taken at a time and place of their choosing, or for a willing physician to administer the drug. The choice would depend on the patient's preference or needs. Before the act is carried out, the patient would also be asked to reiterate their request to ensure full consent.

[151] Fritz (n 20).

The potential problem with vesting this power solely in physicians is that some physicians may not be comfortable with making a positive decision to end a life. Furthermore, there is a risk of a slippery slope effect as a result of busy physicians failing to make a complete and thorough assessment of patients requesting PAD. This could result in on-demand PAD by the back door. Despite these weaknesses, having physicians consider PAD applications by themselves is preferable to handing first instance responsibility to the courts. Doctors have the professional expertise necessary to assess the patient's physical and mental condition and would therefore be capable of assessing whether a patient meets the other conditions. Furthermore, this would be the most cost-effective and efficient option as the decision-making power PAD applications would be vested directly with doctors, requiring no input from independent third parties.

Arms-Length Authority

A third option would be for the Government to set up an independent arms-length authority which would be granted the decision-making power over PAD applications. The UK already has institutional precedent for establishing such an authority effectively in the Human Fertilisation and Embryology Authority (HFEA). This therefore seems a feasible option. The authority could be given the power to licence hospitals and clinics to offer PAD services as well as the duty to review and approve PAD applications by patients. Furthermore, the authority could issue guidance to hospitals and clinics, similar to that issued by the HFEA, on how to ensure that they comply with the law when carrying out PAD procedures.[152]

This option would reduce the pressure on physicians and avoid burdening the court system whilst also reducing the risks of abuse or negligence. Leadership of such an authority would consist of experts with specific knowledge surrounding PAD, including physicians and palliative care practitioners, judges, medical ethicists, and psychologists, as well as 'lay members' with knowledge in areas such

[152] 'How we regulate' (*Human Fertilisation & Embryology Authority*) <www.hfea.gov.uk/about-us/how-we-regulate/> accessed 3 May 2022.

as religion, finance, journalism and the public sector.[153] Although this undeniably makes such an authority the most qualified in making decisions regarding PAD applications, it would also require time and resources to establish and run. Once set up, however, such an authority would be more cost-effective and efficient than relying on the courts and it would be more transparent than vesting decision-making power directly with physicians.

Conclusion

This article has offered a proposal for effectively decriminalising physician-assisted dying in the UK by determining the conditions and safeguards that would be necessary in order to reduce the risks of a slippery slope effect and abuse of vulnerable patients.

Although the Assisted Dying Bill includes some of the conditions and safeguards that have been proposed in this article, others are omitted. These should be included into any future effort to decriminalise PAD in the UK. The key changes which should be made to the Assisted Dying Bill as it stands are:

(1) decriminalisation should extend to active voluntary euthanasia (AVE) as well as physician-assisted suicide (PAS) in order to provide a remedy to patients who are incapable of carrying out the final act themselves;
(2) access should not be limited to terminal patients, but rather it should be granted on the basis of a patient's condition being incurable and irremediable and their suffering being unbearable; and
(3) the introduction of an additional safeguard to counter the increased scope. Namely, a prohibition on physicians starting the conversation surrounding PAD or suggesting it as an option to patients. This would reduce the risk of external pressures influencing a patient's decision.

This article first established the reasons why there is a need for a change in the current law, discussing the issues currently faced by

[153] 'Our people' (*Human Fertilisation & Embryology Authority*) <www.hfea.gov.uk/about-us/our-people/> accessed 3 May 2022.

patients wanting medical assistance to die and documenting the shift in public opinion towards the decriminalisation of PAD.

It discussed the necessary eligibility criteria and threshold conditions, determining that a patient's condition would need to be incurable and irremediable; a minimum age of eighteen; that the patient is experiencing unbearable pain and suffering; and that they have given informed and proper consent to PAD, whether or not they have lost capacity by the time the act is carried out.

It went on to consider various safeguards for both patients and physicians. It proposed necessary safeguards for patients, namely that they would need to have mental capacity and that they would be making an autonomous decision to die, giving fully informed and reiterated consent. It also proposed a fourteen-day cooling-off period between the initial request and fulfilment to ensure that the patient's decision was well-considered. The proposed safeguards for physicians were that two physicians have to be in agreement regarding a patient's condition and their eligibility for PAD, and that physicians must be afforded the right to conscientious objection.

Finally, this article discussed the different options for practical implementation of any future decriminalisation of PAD, determining that the best option would be to vest decision-making power with a new, independent arms-length authority.

Defending the Abused: A Critical Examination of the Defences Afforded to Victims of Domestic Abuse Who Kill

LEWIS WALKER*

Abstract

Murder is often labelled as the most heinous crime. However, the law provides defences for circumstances in which it is believed that the killer's actions were justified or should at least be excused. These defences were constructed to cater to a society dominated by male actions. Therefore, they fail to accommodate for victims of domestic abuse – a category mainly comprised of women – who kill their abusers. This article outlines the defences available in these cases, disposing of self-defence and provocation with ease before critically evaluating the false saviour of diminished responsibility in depth. In doing so, it points towards the need for substantial reform so that the criminal law treats victims of abuse, who end up in its hands, with compassion rather than vilification.

Keywords: domestic abuse, murder, diminished responsibility, provocation, self-defence, coercive and controlling behaviour

Introduction

When victims of domestic abuse are accused of killing their abusers, the law in Scotland lacks compassion.[1] This article discusses the defences available to accused victims of domestic abuse (AVDAs), specifically in cases where they kill their abuser who is in a passive state.[2] Such passive states can be illustrated generally as situations in which the abuser presents no 'immediate' danger to the abused; this

* Law with English Law, LLB (Hons, First Class), University of Aberdeen.

[1] Scotland recently enacted the Domestic Abuse (Scotland) Act 2018, which provides protections to victims of domestic abuse and clamps down on abusers. However, it is silent on the discussion held within this study.

[2] For example, where the abuser is sleeping.

being best exemplified where the abuser is sleeping at the time of the killing. At a basic level of analysis, these killings are most often carried out in defence of the AVDA's own denied autonomy, rather than out of evil intent. On a moral level, it is difficult to class these killings as murder, or at least to do so while showing compassion towards AVDAs, who 'have spent years, sometimes decades, in a prison of their abuse.'[3] Thus the essence of this piece is that, as Dressler puts it, 'we do not want women to kill their sleeping husbands, but we will excuse them under extreme circumstances.'[4]

The law's approach to this problem has been a combination of wilful ignorance and lacklustre engagement. In cases involving ADVAs, the three defences to murder – self-defence, provocation, and diminished responsibility – have been employed to mitigate the ADVA's culpability. However, the use of these has, to understate its failings, been akin to fitting square pegs into round holes. The source of these missteps stems from the roots of the defences in law. Derived from a male-dominated legal system and culture, they are primarily catered to the behaviour of males. In particular, this renders provocation and self-defence inadequate for AVDA-type cases. This is explored in more detail in the first section of this article. The second section then addresses what will be identified as the false promises of diminished responsibility. This section concludes that although at a surface level it may produce adequate results, it causes a greater underlying harm through its 'sydromisation' of women.[5] Thus, an alternative defence drawing from UK legislation and a wide interpretation of common law coercion is suggested.

[3] L Ellison, 'Coercive and Controlling Men and the Women Who Kill Them' (2019) 3 Wolverhampton LJ 22, 35.

[4] J Dressler, 'Battered Women and Sleeping Abusers: Some Reflection' (2006) 3 Ohio State Journal of Criminal Law 457, 470.

[5] P Ahluwalia & A Rafferty, 'Defences available for women defendants who are victims/survivors of domestic abuse' (*Criminal Bar Association of England and Wales*, 17 October 2017) <www.prisonreformtrust.org.uk/portals/0/documents/cba%20domestic%20violence%20briefing.pdf> accessed 15 May 2022.

Excusing Men and Persecuting Women

Before moving forward, the relative gendered nature of domestic abuse should be touched upon. Domestic abuse is framed as a crime indiscriminate as to gender. However, 80% of the recorded incidents in Scotland in 2020/21 involved a female victim and male abuser.[6] Similarly, in England, data shows that women are far more likely to suffer from domestic abuse than men.[7] Therefore, the idea that '[f]emale homicide is so different from male homicide that women and men may be said to live in two different cultures, each with its own 'subculture of violence','[8] has to be understood. The justice system is ingrained to an unbalanced degree with masculinity, which hinders its interactions with female-orientated circumstances.[9] For instance, self-defence and provocation are the offspring of a system catered to male behaviour; called to answer situations involving male-on-male, or male-committed, violence.[10] It follows that females, who make up the majority of VDAs, carry an invisible burden imposed by the law. As AVDAs, they face unfair and unjust labelling, in addition to attempts to rationalise their behaviour as some sort of abnormality. The law rarely views them in the same state it does males. However, these concerns also apply to male VDAs, who may not conform to the stereotypical male image.

The law has almost placed an inferiority tag onto females. Exemplifying this is the historic idea, at least in England, that mariticide without defence was considered petty treason, not the

---

[6] Scottish Government, *Domestic Abuse Recorded by the Police in Scotland, 2018-19* (Scottish Government 2020) 2.

[7] 'Domestic Abuse, Ethnicity Facts and Figures' (*UK Government*, 26 February 2021) <www.ethnicity-facts-figures.service.gov.uk/crime-justice-and-the-law/crime-and-reoffending/domestic-abuse/latest#by-ethnicity-and-sex> accessed 15 May 2022.

[8] L J Taylor, 'Provoked Reason in Men and Women: Heat of Passion Manslaughter and Imperfect Self-Defence' (1986) 33 UCLA L Rev 1679, 1681.

[9] B A Hocking, 'Limited (and Gendered?) Concessions to Human Fragility: Frightened Women, Angry Men and the Law of Provocation' (1999) 6 Psychiatry, Psychology and Law 57.

[10] In relation to self-defence: C Gillespie, *Justifiable Homicide: Battered Women, Self-Defense, and the Law* (Ohio State University Press 1989) 41. For provocation: P Ferguson and C McDiarmid, *Scots Criminal Law: A Critical Analysis* (2nd Edn, Edinburgh UP 2014) 571.

lesser charge of murder.[11] Contrastingly, uxoricide without defence attracted the latter charge. Thankfully, this specific distinction has since been abolished. However, cases involving a female accused are still perceived as underlyingly 'supplementary to the main [male] agenda.'[12] The defences of provocation and self-defence do well to illustrate this when the outcomes of cases involving a female accused are contrasted with those involving males.

Provocation

The English case of *Thornton* is a good starting point.[13] Sara Thornton suffered from a personality disorder and was the subject of repeated violent domestic abuse from her husband.[14] Sara inflicted a fatal blow to her husband with a knife, intending only to scare him after he threatened to kill her in her sleep. Pleas of provocation as well as diminished responsibility received no sympathy from the court,[15] resulting in a life sentence for murder. Historically, provocation was a defence linked to a man's honour.[16] It was born out of misplaced judicial pity towards male responses to sudden emotional upheavals.[17] Take *HM Advocate v Hill*,[18] in which the accused killed his wife and her lover after the former admitted her infidelity. A plea of provocation was accepted, as it was perceived immoral to 'hang a soldier who returned from wars to discover that his wife had been unfaithful,'[19] thus allowing Hill to avoid a conviction for murder.

80 years have passed since *Hill,* but provocation has not deviated from its roots,[20] and it remains suitable (however distasteful

---

[11] W Blackstone, *Commentaries on the Laws of England* (Clarendon Press 1765) 445.
[12] C Wells, 'Provocation the Case for Abolition' in M Ashworth (eds), *Re-thinking English Homicide Law* (OUP 2000) 91.
[13] [1992] 1 All ER 306; [1993] 96 Cr App R 112.
[14] ibid 112.
[15] ibid 118.
[16] J Horder, *Provocation and Responsibility* (OUP 1992) 46.
[17] W Wilson, *Criminal Law Longman Law Series* (7th Edn, Pearson Education Ltd 2019) 394.
[18] 1941 JC 59.
[19] G H Gordon Commentary of *Drury v HM Advocate* 2001 SCCR 583, 618.
[20] Chiu writes that cultures usually end long before their legal doctrines perish. E M Chiu, 'Culture in Our Midst' (2006) 17 University of Florida Journal of Law and Public Policy 231, 245.

one views its existence) primarily for men who lash out in anger.[21] The current state of provocation impresses upon observers that the male outburst is somewhat excusable.[22] Moreover, in cases of domestic abuse, where the male abuser kills their female partner, judges have expressed sympathy towards the male. For example, in the case of Joseph McGrail, Popplewell J stated that McGrail's wife who had been kicked to death by McGrail 'would have tried the patience of a saint,'[23] handing down a two-year suspended sentence. Contrasting these comments and the outcome of *Hill* with *Thornton* does well to expose the gendered standards of the law. Worryingly, the law shows less compassion to AVDAs who have suffered a great deal more both physically and mentally; a man can kill and be excused so long as his patience was tried, but the same leniency is not afforded to abused women who kill their abusers.

Self-defence

Self-defence falls into the same pitfall: it is refined in such a way that justifies primarily male behaviour. There is a general reluctance to expand its definition beyond its traditional application,[24] leaving it too narrow to allow AVDAs who kill their passive abusers to rely on its success.[25] Of course, where the VDA kills her abuser in the midst of violent attack upon herself, self-defence would be the common-sense response to the charge. Yet, the case of Janet Gardner shows that common sense may not be all that common.[26] Although freed from a prison sentence on the basis of provocation and mental disorder, the case is peculiar as Gardner killed her abuser whilst he was strangling her and beating her head against a doorway, but self-defence was not

[21] Ferguson and McDiarmid (n 10) 584.
[22] Horder (n 16) 194.
[23] A McColgan, 'In Defence of Battered Women Who Kill,' (1993) 13 Oxford Journal of Legal Studies 508, 509. Other examples include a case where the judge described a man who had killed his wife as 'a devoted father' who had 'rescued his wife from the gutter.'
[24] ibid 527.
[25] W Chan, 'A Feminist Critique of Self-Defence and Provocation in Battered Women's Cases in England and Wales' (1994) 6 Women and Criminal Justice 39, 43.
[26] H Mills, 'Battered Woman is Freed Over Killing' *Independent* (30 October 1992) <www.independent.co.uk/news/uk/battered-woman-is-freed-over-killing-1560359.html> accessed 15 May 2022.

considered at all by the court.[27] In the passive situations that this article is concerned with, the absence of imminency, proportionality, and necessity, are fatal to pleadings of self-defence by AVDAs. In a situation, say, where the abuser is sleeping, there is no immediate threat to the VDA that renders killing the abuser necessary or proportional. Consequently, many argue for the reform of self-defence to suit the needs of AVDAs.[28] Dressler, however, contends that attacks on the current self-defence are not entirely fair.[29] Although the criminal law is overly steeped in masculine influences,[30] self-defence only looks to preserve human life, nothing else.[31] He remarks that we should altogether approach these reforms reluctantly. We do not wish to over-extend what we consider justifiable deadly force, because we cannot stray from self-defence's purpose to protect all life. Moreover, reforming self-defence is unlikely to solve the same difficulties in instances of pure psychological abuse.[32] Although, Ewing has suggested an interesting reformation of self-defence, based on a psychoanalytical approach.[33] This appears attractive, but its legal and scientific validity are questionable.[34] Diminished responsibility *ex facie* purports to be the best approach, despite it only being a partial defence.[35]

---

[27] McColgan (n 23) 515.

[28] For detailed suggestions for reform, see McColgan (n 23).

[29] J Dressler, 'Feminist (or 'Feminist') Reform of Self-Defense Law: Some Critical Reflections' (2010) 93 Marquette L Rev 1475, 1482.

[30] ibid 1481.

[31] S Estrich, 'Defending Women' (1990) 88 Michigan L Rev 1430, 1431.

[32] However, a reform of self-defence to clarify its definition within the *Draft Criminal Code* should be sought after in general as per Ferguson and McDiarmid (n 10) 567.

[33] C P Ewing, *Battered Women Who Kill: Psychological Self-defence as Legal Justification* (The Free Press 1987) 62. Ewing suggests that the 'self' in self-defence refers to not only our physical bodies but also our psychological functions and attributes. Therefore, we can defend ourselves from 'psychological death' just as we can physical death.

[34] D L Faigman, 'Discerning Battered Women Who Kill' (1987) 39 Hastings LJ 207.

[35] The definition of diminished responsibility can found in the Criminal Procedure (Scotland) Act 1995 s 51B (inserted by The Criminal Justice and Licensing (Scotland) Act 2010): A person who would otherwise be convicted of murder is instead to be convicted of culpable homicide on the grounds of diminished responsibility if the person's ability to determine or control conduct for which the person would otherwise be convicted of murder was, at the time of the conduct,

Diminished Responsibility: A False Solution

Diminished Responsibility in Practice

Diminished responsibility 'poses less of a theoretical puzzle' than provocation.[36] Following *R v Ahluwalia*,[37] diminished responsibility became a go-to defence for battered women who kill their abusers.[38] Any lawyer who disregarded it was perceived as failing in their duties.[39] The link between domestic abuse and psychiatric/psychological disorders is well-documented.[40] Depression, anxiety, substance abuse, and PTSD in women are associated with all forms of abuse.[41] Consequently, the WHO has recognised domestic abuse as a high risk to health.[42] Diminished responsibility latches onto this, contending that VDAs kill due to a mental abnormality caused by long periods of abuse. The practical advantages of this are obvious. With the links to psychological disorders clear, it is easy to accept that an AVDA would kill as a result of the mental abnormality – something jurors are willing to do in order to achieve the least punishable charge.[43] The case of *R v Challen* exemplifies the success of diminished responsibility.[44] *Challen* demonstrated a departure from the regular use of diminished responsibility in these cases in that it failed to mention the much-

substantially impaired by reason of mind. In England, the Homicide Act 1957 s 2 defined it in line with the Scottish common law.

[36] 'Partial Defences to Murder' (2004) Criminal Law Review 1. Consequently, it survived the purge that abolished the English law defence of provocation which in turn was replaced by loss of control.

[37] [1992] 4 All ER 889

[38] V Bettinson, 'Aligning Partial Defences to Murder with the Offence of Coercive and Controlling Behaviour,' (2019) 83 JCL 71, 80.

[39] D Nicholson and R Sanghvi, 'Battered Women and Provocation: The Implication of *R v Ahluwalia*,' (1993) Crim LR 728, 736.

[40] A L Warburton and K M Abel, 'Domestic Violence and Its Impact on Mood Disorder in Women: Implications for Mental Health Workers,' in D Castle and others (eds), *Mood and Anxiety Disorders in Women* (Cambridge UP 2006) 92.

[41] ibid.

[42] 'Violence Against Women' (*World Health Organisation,* 29 November 2017) <www.who.int/news-room/fact-sheets/detail/violence-against-women> accessed 15 May 2022.

[43] M Kasian and others, 'Battered Women Who Kill: Jury Simulation and Legal Defenses' (1993) 17 Law and Human Behavior 289, 310.

[44] [2019] EWCA Crim 916.

maligned Battered Women Syndrome (BWS). In Scotland, diminished responsibility was justified by Battered Woman Syndrome in *Galbraith v HM Advocate (No. 2)*,[45] which placed it upon a pedestal.[46] BWS provides an explanation as to why battered women remain in abusive relationships and/or ultimately end up killing their abuser.[47] In the United States, it provides AVDAs with solutions to evidentiary and jury instruction problems.[48] It utilises the cycle theory which proposes that an abusive relationship cycles through tension, violence, and affection.[49] This causes the VDA to fall into a state of learned helplessness, resulting in a loss of agency.[50] The identification of BWS resulted in one of the only developments in the law that demonstrates the acceptance of AVDAs' violent actions as somewhat reasonable. Thus, in many cases, AVDAs saw their sentences reduced or even fully acquitted.[51]

The False Solution

Wallace notes that the use of BWS made the plight of AVDAs 'judicially cognizable'[52] through a medical diagnosis. Violence by women is often dealt with by means of psychiatric explanations, allowing for the admission of expert evidence.[53] However, as Wells argues in the case against provocation, diminished responsibility and BWS may achieve some protection for abused women, but the secondary consequences are unacceptable.[54] These consequences include the encouragement to pathologise the accused, a key criticism

[45] 2002 JC 1.
[46] R McPherson 'Battered Woman Syndrome, Diminished Responsibility and Women Who Kill: Insights from Scottish Case Law' (2019) 83 JCL 381.
[47] L E Walker, *The Battered Woman Syndrome* (3rd Edn, Springer Publishing Co 2009).
[48] S Wallace, 'Beyond Imminence: Evolving International Law and Battered Women's Right to Self-Defence' (2004) 71 University of Chicago L Rev 1749, 1756.
[49] Walker (n 47).
[50] ibid.
[51] Bettinson (n 38) 385.
[52] Wallace (n 48) 1749. She makes this point in relation to self-defence claims. However, it holds true in relation to diminished responsibility.
[53] C Wells, 'Battered Women and Defences to Homicide: Where Now?' (1994) 14 LS 266, 275.
[54] Wells (n 12) 85.

of BWS.[55] The tendency to use BWS to medicalise VDAs led to Weare, among many, to comment that accused females are dealt with as 'mad, bad or the victim,' rather than reasonable human beings.[56] This is a damning representation of the judiciary; although they can sympathise with male anger, a women needs to be 'defective' to garner the same response. Thus, an AVDA is forced to prove themselves as 'crazy' using BWS, placing their 'personality, characteristics and problems on trial,'[57] and therefore generating disproportionate effects on the female accused.[58]

Moreover, the scientific validity of BWS is easily undermined.[59] Faigman notes that the research shows 'little empirical support for the cycle theory.'[60] He highlights major flaws in the research which push BWS beyond saving.[61] Moreover, he suggests that Walker stretched the learned helplessness theory. Research into this theory showed an inability of those suffering to control the environment around them, thus rendering any hypothesis that abused women could actually kill, a longshot.[62] In summary, BWS has little evidentiary value, despite early commentators seeming to believe it offered a solution to this problem.[63] These flaws lead Faigman to conclude that Walker's work 'is unsound and largely irrelevant to the central issues in such cases.'[64]

---

[55] McPherson (n 46) 386.
[56] S Weare, '"The Mad", "The Bad", "The Victim": Gendered Constructions of Women Who Kill within the Criminal Justice System' (2013) 2 Laws 337.
[57] K O'Donovan, 'Defences for Battered Women Who Kill' (1991) 18 Journal of Law and Society 219, 230
[58] Bettinson (n 38) 80.
[59] ibid.
[60] D Faigman, 'The Battered Woman Syndrome and Self-Defense: A Legal and Empirical Dissent' (1986) 72 Virginia L Rev 619, 636.
[61] M E Seligman and others, 'Alleviation of Learned Helplessness in the Dog' (1968) 73 Journal of Abnormal Psychology 256. The original experiment into learned helplessness involved shocking dogs without presenting them any opportunity to escape. After a while, the dogs readily gave up and did not resist the shocks. When they were later presented with an escape, they did not respond. The dogs only overcame their learned helplessness after being taken from their environment by an external force, not by their own free will. Although, to counter this, one could point to the larger capacity for independence and intelligence in humans.
[62] Faigman (n 60) 641.
[63] ibid 647; see Wallace (n 48).
[64] ibid.

More generally, diminished responsibility suffers from an expansive coverage, something that has not gone unnoticed by its critics.[65] *Galbraith* contributed to this by widening the criteria's catchment.[66] Defence lawyers in claims such as those discussed here, making use of diminished responsibility so to best serve their client in terms of convictions, are likely contributing to this without considering, or simply disregarding, the wider implications. Arguments for the retention of diminished responsibility in general focus on the 'fair and just labelling' of killers who cannot be held fully responsible for their own actions.[67] However, the defence has the opposite effect in the cases relevant to this discussion, and relying too much on such labels could lead to their incarceration on medical grounds.[68] Without adhering to the unfair and unjust labels, AVDAs cannot rely on diminished responsibility. As a result, it is fair to say diminished responsibility approaches defendants with a checklist. Where an AVDA does not tick the 'right' boxes, such as matching the behaviours and mental states associated with BWS, their defence will crumble.[69] Schneider calls this the 'victimization-agency dichotomy.'[70] Where an ADVA does not appear as helpless as the court feels she should be, and exercises some sort of agency, how can they be a victim?[71] Diminished responsibility fails to recognise that the ADVA can be both a victim and an agent.[72] Thus, by relying on diminished responsibility, we are simply allowing the harmful, gendered status quo to live on.[73] Moreover, as noted before, it is only a partial defence. When an AVDA seeks to rely on it, she cannot seek to vindicate herself completely through justification or excuse.[74] Diminished responsibility is a false saviour – perhaps even a wolf in sheep's clothing – and it should be disregarded by AVDAs, despite the ostensibly favourable outcomes of *Ahluwalia* and *Galbraith.*

---

[65] Partial Defences to Murder (n 36).
[66] *Galbraith* (n 45) para 54 per L JG Rodger.
[67] Law Commission, Partial Defences to Murder Final Report (Law Com No 290, 2004), para 5.18.
[68] O'Donovan (n 57) 230.
[69] Wallace (n 48) 1756.
[70] E M Schneider, *Battered Women and Feminist Lawmaking* (Yale UP 2002) 120.
[71] ibid.
[72] ibid.
[73] Gillespie (n 10).
[74] O'Donovan (n 57) 229.

Rescuing Coercive Control from Diminished Responsibility

Does the inadequacy of the current available defences mean that there is no ground for dealing with these cases compassionately, whether that be by justifying or excusing the conduct of AVDAs? Thankfully, there appears to be a lifeline provided for in *Challen,* and Section 76 of the Serious Crime Act 2015 (hereinafter, 'Section 76'). The above analysis illustrates why the means of diminished responsibility used to downgrade Sally Challen's conviction should be criticised, despite the favourable outcome. Instead, the lifeline we can extract from *Challen* is the emphasis that is placed on coercive and controlling behaviour of the abuser and its effects on the AVDA. Coercive control is best defined by sociologist Evan Stark, whose formulation was presented to the court in the *Challen* appeal:

> In coercive control, abusers deploy a broad range of non-consensual, non-reciprocal tactics, over an extended period to subjugate or dominate a partner, rather than merely to hurt them physically. Compliance is achieved by making victims afraid and denying basic rights, resources and liberties without which they are to effectively refuse, resist or escape demands that mitigate against their needs.[75]

Ellison calls *Challen,* and the introduction of coercive and controlling behaviour, 'an amalgamation of old and new law.'[76] The new law being that the appeal was advanced partially on the grounds of coercive and controlling behaviour, a specific form of abuse introduced by Section 76, well after her original murder conviction. The persuasive argument for reviewing the conviction was that an 'advance in understanding amounts to fresh evidence in the same way that advancement of science such as DNA can result in the undermining of the safety of the conviction.'[77] Its amalgamation with existing law being that it prompted suggestions that the defences to

[75] E Stark, *Coercive Control: How Men Entrap Women in Personal Life* (OUP 2007) 5; *Challen* (n 44) para 38.
[76] Ellison (n 3) 28.
[77] F Hamilton, '"My Mum Murdered My Dad. It Was His Fault"' *The Times* (17 February 2018) <www.thetimes.co.uk/article/my-mum-murdered-my-dad-it-was-his-fault-0l9hg955q> accessed 15 May 2022.

murder, particularly provocation and diminished responsibility, are interpreted with this in mind.[78] However, aligning the partial defences, specifically diminished responsibility, with the effects of coercive control, would only serve to paper over the cracks.

Rather than viewing coercive behaviour as a tool to use on the canvasses of diminished responsibility and provocation, it would be much more beneficial to introduce it as its own canvas. In terms of diminished responsibility, there is a clash of rationales when coercive control is introduced. In order to be successful, diminished responsibility requires the presence of an abnormality of the mind on the part of the accused at the time of the incident. Therefore, the AVDA would be required to prove that the coercive and controlling behaviour she was subjected to resulted in her experiencing an abnormal mental state.[79] This conflicts with the purpose of legally recognising coercive and controlling behaviour which looks to attribute the actions of AVDAs directly to the actual abuse, rather than their mental state. Thus, we need to be cautious in moving forward with *Challen* – especially in Scotland where our partial defences lag behind their English counterparts – to ensure that only the beneficial aspects are extracted.

Recognising Human Nature

A new defence that accounts for the circumstances of AVDAs is needed. This defence should be built on the presumption that the right to life is unconditional; therefore, any ideas based on forfeiture (the forfeit of one's right to life through their own conduct) are avoided.[80] Moving forward we can improve upon the use of coercive and controlling behaviour from *Challen* and make it the focal point of a novel defence suggested below.

Ahluwalia and Rafferty suggest that the law should look to mirror legislative provisions used to protect trafficking victims.[81] The Modern Slavery Act 2015 (enacted to implement a European directive

[78] Bettinson (n 38) 71.
[79] Something the defence were unable to do in the original trial of Sally Challen.
[80] For an in-depth argument against the forfeiture policy see Dressler (n 29) 1491.
[81] Ahluwalia and Rafferty (n 5) 9.

into UK law)[82] protects such victims from prosecution if they are compelled to commit an offence due to their circumstances.[83] However, it excludes murder from this defence.[84] This hints towards forfeiture, but because the defence is solely focused on the AVDA being compelled to commit the offence, the principle of these provisions appears to fall more in line with Dressler's idea of fair opportunity.[85] Dressler argues that we can excuse an offence (including murder) if the perpetrator was not afforded 'a fair-opportunity to conform her conduct to the law.'[86] It is quite simple to harmonise coercive and controlling behaviour with the notion of fair opportunity. Dressler's no-fair-opportunity defence focuses on the deprivation of the VDA's autonomy that results in her inability to conform her actions to the law. Stark's definition of coercive control mentioned above highlights the subjugation and domination of a partner through the denial of 'basic rights, resources and liberties'; effectively the deprivation of autonomy.[87] His reference to their own inability to escape their situation is, undeniably, a perfect example of the deprivation of fair opportunity. Therefore, where an AVDA is not presented with a fair opportunity to conform their conduct with the law due to coercive circumstances, thus compelling them to act unlawfully, their conduct should be excused. This allows for VDAs to be viewed as strong, dignified, and reasonable,[88] rather than 'mad, bad or the victim.'[89]

The defence itself can be secured on principled footing. Comparatively, it is similar to the definition of duress contained within the American Model Penal Code,[90] in which the accused is said

[82] Directive 2011/36/EU of the European Parliament and of the Council of 5 April 2011 on preventing and combating trafficking in human beings and protecting its victims, and replacing Council Framework Decision 2002/629/JHA, Article 8.
[83] MSA 2015, s 45.
[84] MSA 2015, sch 4.
[85] Dressler (n 4) 469.
[86] ibid.
[87] Stark (n 75) 5.
[88] ibid 389.
[89] Weare (n 56).
[90] American Model Penal Code s 2.09 provides a defence where the accused acts as a result of prior illegal violence against them or the immediate threat of future violence.

to commit a crime due to prior violent acts towards them.[91] This itself mirrors the defence of coercion (not to be confused with coercive and controlling behaviour) in Scots law,[92] although it would be a slight deviation from the standard interpretation.[93] However, it is not an overly radical step to take, and would be within the rules of interpretation. This should be laid down in legislation, drawing from the 2015 Act. Thus, we are left with a defence laid down in legislation with a solid common law background imported from coercion.

Critics may argue that this defence introduces an uncontrollable subjective element, thus allowing for bad faith claims. It would be correct to define elements of this defence as subjective, but otherwise this is an extremely cynical criticism. The defence can be caveated with objective criteria based on what we already know of domestic abuse. Expert opinions and sound normative judgments can be used to ensure the truth is established. Therefore, each case can stand on its merits while adhering to criteria laid down by the law. Of course, in practice, this defence will encounter difficulties much like every other exception in law. However, its application should produce a policy outcome that properly accounts for the interests of AVDAs.

Conclusion

William Blackstone famously wrote that 'it is better that ten guilty persons escape than that one innocent suffer.'[94] Most often quoted as the rationale for the presumption of innocence, I believe we can take Blackstone's ratio further. The notion appeals to our disgust of injustice; the harm inflicted by a false label is unique. Such harms arise not just when we wrongly label those who have not committed a crime, but also when we incorrectly label those have committed a criminal act. Both acts of injustice pay little attention and compassion to those on the receiving end. The above exploration illustrates how

[91] Dressler (n 4) 470.
[92] Ferguson and McDiarmid (n 10) 545. It should be noted that the relationship between coercion and murder is currently unclear in Scots law.
[93] Usually, duress (in American and English Law) and coercion (Scots law) are interpreted as requiring some sort of order from the original violent actor placed upon the accused pleading the defence.
[94] Blackstone (n 11) 358.

our system struggles to overcome its tendency to the former when it comes to AVDAs.

In short, not one of the three defences to murder are fit to channel the compassion AVDAs desperately need. The analysis of provocation does well to illustrate the hole it digs itself as it refuses to remove itself from its gendered history. Self-defence suffers from the same problem of being born from a system of law dominated by men. It was never designed to cater for circumstances that AVDAs find themselves in; where the physical parity between abused and abuser is so insurmountable that the abused resorts to killing their abuser while the latter sleeps. Most unsettling, though, is the current tendency of AVDAs to feel the need to frame themselves as 'psychologically abnormal' in order to successfully plead diminished responsibility. It is undignified to assume that a victim of domestic abuse has killed primarily because she is suffering from psychological harm and not because she is, in fact, a reasonable human being. However, due to its previous successes, diminished responsibility presents a tempting, but dangerous, offering for ADVAs and their defence counsels. To avoid a repeat of the injustices of *Thornton* and the poisonous methods of the diminished responsibility cases, a new defence to murder is needed specifically for ADVA-type cases. Adopting the fair opportunity paired with recognition of coercive and controlling behaviour would 'make law fit for women's experience.'[95] Compassion would overcome injustice, recognition of autonomy would replace insincere pity, and, most importantly, everyone would be given a fair opportunity before the law.

[95] L Finlay, 'Reflections on Feminism Unmodified' (1988) 82 Northwestern L Rev 352, 385.

Parliamentary Sovereignty in Canada: A Divergent Perspective from the United Kingdom

TARKEN JAMES MOORE*

Abstract

The principle of parliamentary sovereignty was directly adopted from the UK into the constitutional order of Canada upon the country's formation as a Dominion of the British Empire in 1867. Departing from the British understanding of the principle, the Canadian and British courts would come to recognise a uniquely Canadian understanding of parliamentary sovereignty. The sovereignty of the Canadian Parliament and the provincial legislatures would gradually grow as the Dominion acquired greater autonomy throughout the early twentieth century, reaching its apex following the enactment of the Statute of Westminster in 1931. This article examines how parliamentary sovereignty was adopted into the Canadian constitutional order in 1867 and its subsequent growth over the following 115 years. It concludes by discussing the principle's curtailment in 1982 following the enactment of the Canadian Charter of Right and Freedoms and the reduced role that it plays in Canada's contemporary constitutional order.

Keywords: Canada, United Kingdom, parliamentary sovereignty, constitutional law

Introduction

The doctrine of parliamentary sovereignty has occupied a leading position in Canada's constitutional order since Canadian Confederation in 1867, when the Fathers of Confederation, Canada's founders, adopted the principle from the United Kingdom in their efforts to unite the colonies of British North America into the Dominion of Canada.[1] As a consequence of Canada's federal system

* LLB (Hons, First Class) University of Aberdeen, 2021; LPC, BPP University, 2022. The author spent over two years working as an advisor to a Canadian MP. He will

of government and adherence to constitutional supremacy, the principle of parliamentary sovereignty has come to be enjoyed not only by the Parliament of Canada, but by the provincial legislatures as well. This divergent application of the principle, along with several other key differences, are the result of the differing natures of the British and Canadian constitutions having forced parliamentary sovereignty's doctrine to adapt in order to function within the Canadian legal order.

The fact that the principle of parliamentary sovereignty, as it has come to be understood in Canada, has managed to successfully depart almost entirely from the orthodox teachings of Dicey whilst maintaining many of its essential features, raises two important questions. First, how has parliamentary sovereignty evolved in Canada since Confederation in 1867? Second, how does it compare to parliamentary sovereignty in the United Kingdom?[2]

This article brings awareness to the manner in which parliamentary sovereignty has transmuted doctrinally in order to insert itself into the constitutional order of Canada. It explores how the Canadian conceptualisation of the principle has evolved since Confederation in light of Canada's growing autonomy from the United Kingdom and eventual adoption of the Canadian Charter of the Rights and Freedoms. In doing so, this article examines how Canadian parliamentary sovereignty has come to differ from its British progenitor, challenging the British perception that parliamentary sovereignty is a rigid principle.

This article is divided into five chapters. The first chapter examines the orthodox doctrine of British parliamentary sovereignty

---

begin studying the LLM in International Business Law at King's College London in 2022.

[1] British North America Act 1867, preamble; Craig Forcese, Adam Dodek; Philip Bryden and others (eds), *Public Law: Cases, Commentary, and Analysis* (3rd edn, Emond Publishing 2015) 155; Norman L Nicholson, 'British North America' (*The Canadian Encyclopedia*, 6 February 2006) <www.thecanadianencyclopedia.ca/en/article/british-north-america> accessed 20 February 2021.

[2] AV Dicey, *Introduction to the Study of the Law of the Constitution* (8th edn, Macmillan 1915) pt 1.

and rationalises the limitations that have come to be recognised by the British courts in the context of Hart's rule of recognition.[3] The second chapter analyses how parliamentary sovereignty was adopted into Canada's constitutional order at Confederation and the adjustments that were needed in order for it function in light of Canada's subordinate relationship to the United Kingdom and federal system of government. The third chapter looks at the major constitutional developments that took place between 1867 and 1982 and the impact that they had on Canadian parliamentary sovereignty. The fourth chapter discusses the impact that patriation had on the sovereignty of Parliament and the provincial legislatures in light of their newly acquired independence and the relationship between the principle and the Canadian constitution's amendment procedure. The fifth chapter examines the impact that the Canadian Charter of Rights of Freedoms had on parliamentary sovereignty and how it greatly reduced the sovereignty of both Parliament and the provincial legislatures.

Parliamentary Sovereignty in the United Kingdom

The Orthodox Doctrine of British Parliamentary Sovereignty

The most widely cited and influential account of British parliamentary sovereignty is that of jurist and constitutional theorist Albert Venn Dicey.[4] Despite its orthodox nature and partial incompatibility with the principle's more recent legal developments and considerations, Dicey's account remains an accurate and reliable expository of the fundamental aspects of the doctrine of parliamentary sovereignty. Thus, it continues to be influential in court proceedings.[5]

In his book, *Introduction to the Study of the Law of the Constitution*, Dicey states that the principle of parliamentary sovereignty encompasses British Parliament's 'right to make or unmake any law' with 'no person or body [being] recognised (…) as having a right to

[3] HLA Hart, *The Concept of Law* (3rd edn, OUP 2012) ch 6.
[4] *Attorney-General (New South Wales) v Trethowan* [1932] AC 526; *R (Miller) v Secretary of State for Exiting the European Union* [2017] UKSC 5.
[5] ibid.

override or set aside the legislation of Parliament.'[6] As Lord Reid explains:

> It is often said that it would be unconstitutional for the United Kingdom Parliament to do certain things, meaning that the moral, political or other reasons against doing them are so strong that most people would regard it as highly improper if Parliament did these things. But that does not mean that it is beyond the power of Parliament to do such things. If Parliament chose to do any of them the courts could not hold the Act of Parliament invalid.[7]

Although Lord Reid sets out the parameters of parliamentary sovereignty, he does not elaborate on its source. According to Dicey, the unbridled authority of Parliament is intrinsic to its state of being sovereign, in which it is free of any external constraints, such as an entrenched constitution or a superior legislature.[8] Austin states that in order for a body to be considered sovereign, it must '*not* [be] in a habit of obedience to a determinate human superior' and its 'power… in its… sovereign capacity' must be 'incapable of legal limitation.'[9] Although Dicey distinguishes his treatment of sovereignty from that of Austin on several points,[10] the fundamental characteristics that define sovereignty are largely alike in both of their accounts.

As a consequence of Parliament's sovereign nature, Dicey asserts that it is incapable of legislating in a manner that would limit the powers of or legally bind either itself or a successor Parliament.[11] While ostensibly contradictory to the idea that Parliament possesses absolute legislative authority, Dicey justifies this assertion on the basis 'that a sovereign power cannot, while retaining its sovereign character, restrict its own powers by any particular enactment.'[12]

---

[6] Dicey (n 2) 3-4.
[7] *Madzimbamuto v Lardner-Burke* [1969] 1 AC 645 (HL) (Lord Reid).
[8] Dicey (n 2) 24-25; Mark Elliott and Robert Thomas, *Public Law* (3rd edn, OUP 2017) 233.
[9] John Austin, *The Province of Jurisprudence Determined* (John Murray 1832) 199, 258.
[10] Dicey (n 2) 26-30.
[11] ibid 21.
[12] ibid 24.

The Limitations of British Parliamentary Sovereignty

Contradicting the orthodox assertions of Dicey, the judiciary of the UK has since come to recognise several limitations on the legislative powers of Westminster over the course of the twentieth and twenty-first centuries. Although these limitations do not directly undermine the parameters of parliamentary sovereignty, their successful implementation demonstrates that the principle's efficacy and legal paramountcy stems not from Parliament's sovereign nature, but instead from court recognition. This understanding of parliamentary sovereignty lies at the core of the principle's emergence and evolution in Canada. It is therefore necessary to examine its theoretical underpinnings and application in the United Kingdom before looking to Canadian parliamentary sovereignty.

The role that court recognition plays in enforcing and limiting parliamentary sovereignty can best be explained using Hart's expository account in *The Concept of Law*. According to Hart, a legal system is a union of primary and secondary rules, with primary rules being those that impose duties and confer power (for example, Acts of Parliament, statutory instruments and case law), and secondary rules being those 'that provide for the identification, alteration, and enforcement of the primary rules'.[13] Further subdividing secondary rules into three categories, Hart posits that there exist rules of recognition: the 'criteria by which the validity of the other rules of [a legal] system is assessed'.[14] Although rarely referred to directly by the courts, Hart explains that the rule of recognition is generally '*shown* in the way in which particular rules are identified' through the process of adjudication.[15] Wade elaborates upon this by explaining that 'the rule of recognition is itself a political fact which the judges themselves are able to change when they are confronted with a new situation which so demands.'[16]

The rule of recognition that was traditionally employed by the courts during Dicey's lifetime could have been phrased, 'an Act of the

[13] Hart (n 3) xix-xx.
[14] ibid 105.
[15] ibid 101 (emphasis added).
[16] HWR Wade, 'Sovereignty – revolution or evolution?' (1996) 112 LQR 568, 574.

Parliament of the United Kingdom is law.'[17] Despite this, the House of Lords would come to rule in *R (Factortame Ltd) v Secretary of State for Transport*[18] that,

> a national court which, in a case before it concerning [EU] law, considers that the sole obstacle which precludes it from granting interim relief is a rule of national law must set aside that rule.[19]

This ruling would amount to a change in the United Kingdom's rule of recognition as it had effectively given EU law primacy over Acts of Parliament in direct contradiction of the orthodox understanding of parliamentary sovereignty. Although the UK has since left the EU, ending the supremacy of new EU law,[20] this is an important limitation to consider.

Subsequently, in the case of *Thoburn v Sunderland City Council*, the Queen's Bench Division would rule that Parliament was unable to implicitly repeal certain statutes, such as the Human Rights Act 1998 and the Scotland Act 1998, on account of their constitutional significance.[21] This marked yet another departure from the Dicean understanding of parliamentary sovereignty: specifically the understanding that there is 'no marked or clear distinction between laws which are not fundamental or constitutional and laws which are fundamental or constitutional'.[22]

The recognition, application, and enforcement of these limitations in subsequent case law vindicated the Hartian perspective that the doctrine of parliamentary sovereignty derives its legitimacy and efficacy from court recognition and refuted the orthodox notion that the principle is rooted in the sovereign nature of Parliament.[23] Although the courts have imposed several limitations on

---

[17] Hart (n 3) xx.
[18] [1991] AC 603.
[19] ibid 644-45.
[20] European Union (Withdrawal) Act 2018, s 1; European Union (Withdrawal Agreement) Act 2020, s 38.
[21] [2002] EWHC 195 (Admin).
[22] Dicey (n 2) 37.
[23] *R (Jackson) v Attorney General* [2006] 1 AC 262, 303.

parliamentary sovereignty through their alteration of the UK's rule of recognition, they have not altered it to the extent so as to impose any 'hard' limitations on the parameters of the principle that would otherwise prevent Parliament from legislating on any given matter.[24] Considering that these limitations can be unilaterally removed[25] or sidestepped[26] with an expressly worded Act of Parliament, they can be considered 'soft' limitations as they do not ultimately preclude Parliament from passing any law it wishes. Nevertheless, it would be through the recognition of similar limitations by the courts that Canadian parliamentary sovereignty would come to develop and evolve over the course of a century and a half.

The Emergence of Canadian Parliamentary Sovereignty

Adopting Parliamentary Sovereignty in the Dominion

On 1 July 1867, the Dominion of Canada was established by Royal Proclamation[27] after years of negotiations between the United Kingdom and the colonies of British North America.[28] Made up of the former colonies of New Brunswick, Nova Scotia, and the Province of Canada, the new Dominion took the form of a federation to alleviate the linguistic, religious and political differences that would have otherwise made Canadian Confederation untenable.[29] To give effect to Confederation and the Dominion's federal system of government, the Parliament of the United Kingdom enacted the British North

[24] Adam Tucker, 'Uncertainty in the Rule of Recognition and in the Doctrine of Parliamentary Sovereignty' (2011) 31 OJLS 61.
[25] For example, European Union (Withdrawal) Act 2018, s 1.
[26] For a relevant pre-*Thoburn* example of Parliament expressly dissolving a devolved legislature see Northern Ireland Constitution Act 1973.
[27] 'A Proclamation For uniting the Provinces of Canada, Nova Scotia, and New Brunswick, into one Dominion under the name of CANADA' *The Canada Gazette* (Ottawa, 1 July 1867); British North America Act 1867, s 3.
[28] Margaret Conrad, *A Concise History of Canada* (CUP 2012) ch 6; PB Waite, 'Confederation' (*The Canadian Encyclopaedia*, 22 September 2013) <www.thecanadianencyclopedia.ca/en/article/confederation> accessed 20 February 2021.
[29] Eugene A Forsey, *How Canadians Govern Themselves* (10th edn, Library of Parliament 2020) 7-8.

America Act 1867. This was the first of several British North America Acts that would make up the Constitution of Canada.[30]

Under the preamble to the British North America Act 1867, Canada's founding colonies expressed a collective desire to adopt 'a Constitution similar in principle to that of the United Kingdom' for the new Dominion.[31] This resulted the adoption of the principles and conventions of the contemporary British constitutional order into that of Canada.[32] Anticipating the limitations that would be imposed on these principles by Canada's subordinate position to the United Kingdom and partially codified constitution, the Fathers of Confederation issued the following resolution at the Québec Conference of 1864,[33] clarifying how they intended for such principles to operate in light of Canada's circumstances:

> In framing a Constitution for the General Government; the Conference, with a view to the perpetuation of our connection with the Mother Country, and to the promotion of the best interests of the people of these Provinces, desire to follow the model of the British Constitution *so far as our circumstances will permit*.[34]

Although the Fathers of Confederation made no mention of parliamentary sovereignty, the principle would come to be recognised as an integral part of the Constitution of Canada by the courts as a direct consequence of adopting British constitutionalism in Canada. However, several 'fundamental departures' from the British

[30] The British North America Acts were renamed the Constitution Acts in 1982 by the Canada Act 1982, s 53. Due to its historical nature, this article will use the original terminology.
[31] British North America Act 1867, preamble.
[32] Forcese and others (n 1) 155.
[33] Andrew McIntosh and Phillip A Buckner, 'Quebec Conference, 1864' (*The Canadian Encyclopaedia*, 7 August 2006) <www.thecanadianencyclopedia.ca/en/article/quebec-conference> accessed 23 February 2021.
[34] 'Report of Resolutions Adopted at a Conference of Delegates From The Provinces of Canada, Nova Scotia And New Brunswick, And The Colonies of Newfoundland And Prince Edward Island, Held at The City of Quebec, 10th October, 1864, as The Basis of a Proposed Confederation of Those Provinces And Colonies' [1865] Journal of the House of Assembly 854, 854 (emphasis added).

understanding of parliamentary sovereignty were necessary to ensure conformity with Canada's constitutional order. This meant the Canadian principle would differ significantly from its British forerunner.[35] Canadian parliamentary sovereignty is therefore best understood as its own distinct principle, separate from that of the United Kingdom. Furthermore, Canadian parliamentary sovereignty cannot be examined from a Dicean perspective for it neither conforms with orthodox British accounts nor does it emanate from the sovereignty of Canada's Parliament or legislatures. Instead, Canadian parliamentary sovereignty must be considered from a Hartian perspective in that its existence and position in Canada's constitution is rooted in its recognition by the courts.

Colonial Laws Validity Act 1865, sections 3 and 4

The main statute giving effect to the Canadian conceptualisation of parliamentary sovereignty was the Colonial Laws Validity Act (CLVA) 1865, which had been enacted by Westminster prior to Confederation. According to Bird,

> [the] CLVA [was] concerned with validating colonial laws. More precisely, the CLVA validated laws enacted by British colonies to the extent that these laws were consistent with British laws that applied to the colonies'[36]

Although colonial legislatures had possessed the authority to legislate prior to 1865, the CLVA 'abrogated the old rule requiring colonial enactment to conform as near as possible with the fundamental principles of English law'.[37]

The primary reason for the CLVA's enactment was to address the issue of judges 'invalidating colonial laws because of their inconsistency with British laws without regard to whether the British

[35] Peter W Hogg, *2012 Student Edition: Constitutional Law of Canada* (Carswell 2012) 12-4.
[36] Brian Bird, 'The Unbroken Supremacy of the Canadian Constitution' (2018) 55 Alberta Law Rev 755, 757.
[37] Enid Campbell, 'Colonial Legislation and The Laws of England' (1965) 2 University of Tasmania L Rev 148, 148.

laws in question were intended to apply to the colony.'[38] The most notorious example of this was in the Colony of South Australia, which for the decade leading up to the Act had seen Boothby J make rulings that went far beyond requiring compliance with UK law.[39] In *Liebelt v Hunt*, he went far as to rule in that the Parliament of South Australia lacked the authority to depart with the English common law.[40] The Colony of South Australia's situation was so severe that the CLVA had to dedicate an entire provision to retrospectively validating the colonial Parliament's laws.[41]

For the rest of the colonies, section 3 of the CLVA provided that:

> No colonial law shall be or be deemed to have been void or inoperative on the ground of repugnancy to the law of England, unless the same shall be repugnant to the provisions of some such Act of Parliament, order, or regulation as aforesaid.

Section 4 provided for the validity of colonial laws that were inconsistent with the instructions given to colonial governors. Accordingly, sections 3 and 4 of the CLVA had the effect of making Acts of the Canadian Parliament and provincial legislatures non-justiciable to the extent that they were consistent with the British North America Acts and any other UK legislation that was directly applicable to Canada. This prevented British judges from using the common law or any non-applicable legislation as grounds for invalidating their enactments. This would effectively render Parliament and provincial legislatures jointly 'sovereign' over all matters that had not been reserved by Westminster. In the words of AV Dicey, the Parliament and provincial legislatures of Canada possessed the 'right to make or unmake any law' within their *respective scopes* of authority with 'no person or body [being] recognised (…) as

[38] Bird (n 36) 757.
[39] ibid.
[40] *Liebelt v Hunt* (Supreme Court of South Australia, 13 June 1861).
[41] Colonial Laws Validity Act 1865, s 7.

having a right to override or set aside [their] legislation' falling within these scopes aside from Westminster itself.[42]

Departing with the British Conceptualisation of Sovereignty

As discussed, modern accounts such as Hart diverge from Dicey on the source of British parliamentary sovereignty. However, they agree on the fact that Parliament itself is a sovereign law-making body in the sense that it is not subordinate to any other person or body. Although Westminster had initially acquired its unqualified sovereign nature from the Bill of Rights in 1689, it is no longer reliant upon this enactment to give effect to its sovereignty.[43] Were Westminster to repeal the Bill of Rights tomorrow, this would have no bearing on its sovereignty, as it would not result in its subordination to an external authority.

On the contrary, the 'sovereignty' of the Parliament and legislatures of Canada were both conditional and restricted to certain legislative matters at Confederation. Parliament's sovereignty was reliant on an imperial statute. Canada was afforded near absolute autonomy over all of its internal affairs upon its founding.[44] However, if desired, Westminster could have disapplied sections 3 and 4 of the CLVA or repealed the British North America Act 1867, putting an end to the Canadian Parliament's 'sovereignty'. Dicey specifically addressed this issue, asserting that:

> because the colonies, however large their practical freedom of action, do not act as independent powers in relation to foreign states; the Parliament of a dependency cannot itself be a sovereign body.'[45]

It cannot be denied that the Parliament and provincial legislatures of Canada were subordinate law-making bodies to

[42] Dicey (n 2) 3-4.
[43] Bill of Rights [1688]. See Jeffrey Goldsworthy, *The Sovereignty of Parliament: History and Philosophy* (OUP 2001), ch 7.
[44] Brian M Mazer, 'Sovereignty and Canada: An Examination of Canadian Sovereignty from a Legal Perspective' (1977) 42 Saskatchewan L Rev 1, 11.
[45] Dicey (n 2) 61.

Westminster at Confederation. However, in the context of this relationship, Canadian 'sovereignty' was not purporting to be anything other than non-justiciable autonomy and authority over matters falling within the legislative scopes of Parliament and the legislatures. The principle of parliamentary sovereignty had effect to the extent that this was possible in Canada's circumstances.

Colonial Laws Validity Act 1865, section 2

Sections 3 and 4 of the CLVA effectively rendered the Canadian Parliament and provincial legislatures 'sovereign' over all matters that had not been reserved by Westminster. Section 2 of the CLVA expressly restricts Canadian parliamentary sovereignty. It provided for the voidance of Dominion laws that were

> repugnant to the Provisions of any Act of Parliament extending to the [Dominion] to which such Law may relate, or repugnant to any Order or Regulation made under Authority of such Act of Parliament, or having in the [Dominion] the Force and Effect of such Act.

According to Bird, section 2 of the CLVA functioned as a *de facto* supremacy clause for the British North America Acts and applicable UK legislation. It had the effect of barring the Dominion from enacting legislation that was inconsistent with their provisions.[46] However, the courts rarely cited the CLVA in their rulings. Barry L Strayer (former federal Assistant Deputy Minister of Justice) explains that

> [when] a Canadian court struck down a statute for constitutional invalidity, it was inarticulately applying the [CLVA], holding void the Canadian statute for repugnancy to the provisions of the [British North America Acts] distributing power between Parliament and Legislatures.'[47]

[46] Bird (n 36) 756-59.

[47] Barry L Strayer, *The Canadian Constitution and the Courts: The Function and Scope of Judicial Review* (3rd edn, Butterworths 1988) 7.

This was affirmed by the Supreme Court of Canada in *Reference Re Manitoba Language Rights* over one hundred and eighteen years following Confederation.[48]

That said, section 2 of the CLVA only voided a Dominion law to the extent of its repugnancy with an applicable UK law. It did not go any further than it needed to, respecting the sovereignty of the Canadian Parliament and legislatures over the matters falling within their jurisdictions.[49] It can therefore be said that section 2 of the CLVA established the boundaries of Canadian parliamentary sovereignty until the enactment of the Statute of Westminster in 1931. Provided they legislated within their respective scopes of authority, Parliament and provincial legislatures were sovereign. However, the scope of their sovereignty was limited by section 2.

Applying the writings of Hart, Canada's rule of recognition at Confederation could be stated as: 'the Parliament and provincial legislatures of Canada shall be 'sovereign' to the extent that they legislate consistently with the British North America Acts and any other directly applicable UK legislation per section 2 of the CLVA'.

Recognising Canadian Parliamentary Sovereignty in Case Law

Just as the courts gave effect to the sovereignty of the Westminster Parliament following the enactment of the Bill of Rights,[50] so the Canadian courts gave effect to Canadian parliamentary sovereignty under the common law following Confederation. According to Kazmierski, judges would 'emphasis[e] the near absolute authority provided by the principle of parliamentary sovereignty in those areas that the [written constitution] doesn't apply'. This was in keeping with Canada's inherited tradition of uncodified constitutionalism whilst simultaneously respecting its codified constitution.[51]

[48] (1985) 1 SCR 721, 746.
[49] Colonial Laws Validity Act 1865, s 2.
[50] Goldsworthy (n 43) chs 7-8.
[51] Vincent Kazmierski, 'Draconian but Not Despotic: The Unwritten Limits of Parliamentary Sovereignty in Canada' (2009) 41(2) Ottawa Law Review 245, 250.

The first significant case in which the principle of Canadian parliamentary sovereignty was recognised by the judiciary was *Hodge v R*.[52] In *Hodge*, the Judicial Committee of the Privy Council (the House of Lords) of the United Kingdom held that:

> [the British North America Act 1867] conferred powers not in any sense to be exercised by delegation from or as agents of the Imperial Parliament, but authority as plenary and as ample within the limits prescribed by section 92 as the Imperial Parliament in the plenitude of its power possessed and could bestow. Within these limits of subjects and area the local legislature is supreme, and has the same authority as the Imperial Parliament, or the Parliament of the Dominion, would have had under like circumstances.[53]

This ruling amounted to judicial recognition from the UK's highest court that the law-making powers of the Parliament and provincial legislatures of Canada were greater than a mere delegation of legislative authority (i.e. greater than devolution). In spite of the imperial relationship subsisting between the Dominion and the mother country, once competently enacted, the laws of the Canadian Parliament and legislatures were equal in status to laws passed by Westminster. While the House of Lords did not actually use the word 'sovereign', it did recognise that the Parliament and provincial legislatures of Canada are 'sovereign' within their respective limits under the British North America Act 1867. This gave the concept of Canadian parliamentary sovereignty significant doctrinal weight in Canada's constitutional order. Indeed, it would come to be recognised as the paramount uncodified principle of the Canadian constitution.[54]

In *Florence Mining Co Ltd v Cobalt Lake Mining Co Ltd*, Ridell J would elaborate on the nature of Canadian parliamentary sovereignty that had come to be recognised by the judiciary. He explained, in the context of the provincial legislatures, that:

---

[52] [1883] 9 AC 117; Maxime Charron-Tousignant and Robin MacKay, *Parliament and the Courts: Balancing the Roles* (Library of Parliament 2015).
[53] [1883] 9 AC 117, 132.
[54] Hogg (n 35) ch 12; *Bacon v Saskatchewan Crop Insurance Corp* (1990) 180 SaskR 20 [29].

> In short, the Legislature within its jurisdiction can do everything that is not naturally impossible, and is restrained by no rule human or divine. If it be that the plaintiffs acquired any rights, which I am far from finding, the Legislature had the power to take them away. The prohibition, "Thou shall not steal," has no legal force upon the sovereign body. And there would be no necessity for compensation to be given. We have no such restriction upon the power of the Legislature as is found in some States.[55]

This passage strongly mirrors that of Lord Reid in *Madzimbamuto v Lardner-Burke*.[56] It illustrates that the Canadian Parliament and the provincial legislatures are not bound by moral or political considerations when exercising their authority within their respective jurisdictions and thus possess absolute legislative discretion.

Federalism's Role in Shaping Canadian Parliamentary Sovereignty

Parliamentary sovereignty was recognised as the paramount uncodified principle in the Canadian constitution at Confederation. However, it had to significantly adapt to the codified principle of federalism that was embedded in the British North America Act 1867, which was the primary basis for the Act in the first place.[57] Had Canada not structured itself as a federation, it is likely that it would have had an uncodified constitution similar to New Zealand, which actually observes a more 'pure' version of parliamentary sovereignty than the United Kingdom does in the present day.[58]

Sections 91 and 92 of the British North America Act 1867 enumerate the heads of power that fall exclusively within the jurisdictions of Parliament and the provincial legislatures respectively.[59] The House of Lords explained in *Hodge v R* that

---

[55] (1909) 18 OLR 275, 279 (Ridell J).
[56] [1969] 1 AC 645, 723 (Lord Reid).
[57] Forcese and others (n 1) 155.
[58] Karen Grau, 'Parliamentary Sovereignty: New Zealand - New Millennium' [2002] Victoria University of Washington L Rev 12; Hogg (n 35) 12-3.
[59] Forcese and others (n 1) 159.

> When the British North America Act enacted that there should be a legislature for Ontario, and that its legislative assembly should have exclusive authority to make laws for the Province and for provincial purposes in relation to the matters enumerated in section 92… and [the legislature] has the same authority as the Imperial Parliament, or the Parliament of the Dominion, would have had under like circumstances ...[60]

Applying the same reasoning to the Parliament of Canada in relation to section 91, it can be understood that the Parliament and provincial legislatures of Canada are considered to be *exclusively sovereign* over the matters falling within their respective scopes of authority. As Hartery explains, '[n]either order of government is subordinate to the other. They are instead coordinate, both being sovereign'.[61] This differs significantly from the British understanding of parliamentary sovereignty, which considers sovereignty to be vested indivisibly in a single omnipotent law-making body that is free from any 'hard' external restrictions.[62]

Dicey would personally reject the notion that federalism and parliamentary sovereignty were compatible in any way, asserting that

> [e]very legislative assembly existing under a federal constitution is merely… a subordinate law-making body, whose laws are of the nature of bye-laws, valid whilst within the authority conferred upon it by the constitution.[63]

His reasoning for this was that a division of powers under a codified constitution restricted the law-making powers of all legislative bodies, rendering them subordinate to the constitution.[64] While modern accounts of British parliamentary sovereignty do not put so much emphasis on a law-making body's sovereign nature, they

[60] [1883] 9 AC 117, 132.
[61] Jesse Hartery, 'Protecting Parliamentary Sovereignty' (2020) 58 Atlanta Law Rev 187, 189.
[62] Dicey (n 2) ch 3.
[63] ibid 82.
[64] ibid ch 3.

have not come to recognise any permanent limitation as being compatible with the principle.[65]

Although it cannot be denied that there is an 'internal contradiction in speaking of federalism in the light of the invariable principle of British parliamentary supremacy',[66] Canadian parliamentary sovereignty has nevertheless nestled surprisingly neatly into a federal system that treats sovereignty as a divisible concept. As Hogg explains, neither Parliament nor the legislatures are sovereign in the sense of being able to make or unmake any law whatsoever but are instead 'sovereign' within their respective scopes of authority.[67] Considering Canada's constitutional circumstances, an understanding like this is still quite remarkable. Most federal systems (for example, the United States and Germany) do not consider their legislatures to be sovereign in any sense. This is indicative of the great lengths that the Canadian legal system has gone in replicating the British constitution in the Dominion.

What was the CLVA's role in respect to federalism at Confederation? Although the CLVA shielded the Canadian Parliament and legislatures from judicial scrutiny when they were legislating within their respective heads of power,[68] it exposed them to judicial review if they were thought to have exceeded that boundary.[69] *Prima facie* this would appear to violate the parameters of parliamentary sovereignty; that 'no person or body is recognised… as having a right to override or set aside [their] legislation'.[70] However, when considered within the Canadian federal system, the CLVA can be understood to be preventing the Canadian Parliament and the provincial legislatures from interfering with one another's affairs. Thus, the 'sovereignty' of both Parliament and the provincial legislatures and their right to make or unmake any law within their respective scopes of authority was preserved. It can therefore be said that invalidating legislation on the basis of federalism in Canada does

[65] Some will argue that the Treaty of Union is binding, although this is not widely accepted. See *MacCormick v Lord Advocate* 1953 SC 396, 412 (Lord Cooper)
[66] *Reference Re Resolution to amend the Constitution* (1981) 1 SCR 753, 806.
[67] Hogg (n 35) 12-14.
[68] Colonial Laws Validity Act 1895, ss 3-4.
[69] ibid s 2.
[70] Dicey (n 2) 38.

not restrict the 'sovereignty' of the offending legislative body, but instead upholds the 'sovereignty' of the other body by ensuring that its legislation is not overridden or set aside.

Although sections 91 and 92 of the British North America Act 1867 expressly listed out a wide range of matters within the competences of the Canadian Parliament and provincial legislatures, these lists were by no means exhaustive. Legislatures subject to a codified constitution cannot normally create 'a new legislative power not created by the Act to which it owes its own existence'.[71] On the other hand, Canada's doctrine of exhaustion allows both Parliament and the provincial legislatures to legislate on matters not mentioned in sections 91 or 92 provided that they legislate consistently with the British North America Acts and any applicable UK legislation.[72] However, this is subject to the doctrine of paramountcy, as set out by Lord Dunedin:

> First, that there can be a domain in which provincial and Dominion legislation may overlap, in which case neither legislation will be ultra vires, if the field is clear; and, secondly, that if the field is not clear, and in such a domain the two legislations meet, then the Dominion legislation must prevail.[73]

Federal laws are given precedence over provincial laws on matters falling outwith sections 91 and 92 of the British North America Act 1867. Although the provincial legislators have the conditional right to legislate in these areas, only Parliament can be understood to be 'sovereign' over them.

That said, although Canada adopted a federal system of government at Confederation, the Dominion would not observe any formal division of powers between its legislative and executive branches. It had adopted the British parliamentary system.[74] On

---

[71] *Re the Initiative and Referendum Act* (1919) 48 DLR 18, 25.
[72] Forcese and others (n 1) 155.
[73] *Grand Trunk Railway Company of Canada v Attorney General of Canada* [1907] AC 65, 67 (Lord Dunedin).
[74] Wan Wai Yee, 'Recent Changes to the Westminster System of Government and Government Accountability' (1994) 15 Singapore L Rev 297; British North America Act 1867, pts 3-4.

account of this, there was nothing that could be done by the federal or provincial cabinets that could not be done either by Parliament or the provincial legislatures. The executive branch was therefore entirely subordinate to the sovereignty of the legislative bodies.[75] The Supreme Court of Canada would elaborate on this relationship in the context of parliamentary sovereignty in *Canada (Auditor General) v Canada (Minister of Energy, Mines and Resources)*, explaining that

> [t]he *grundnorm* [the highest authority] with which the courts must work in this context is that of the sovereignty of Parliament. The ministers of the Crown hold office with the grace of the House of Commons and any position taken by the majority must be taken to reflect the sovereign will of Parliament.[76]

Major Developments between 1867 and 1982

Acquiring and Administering the Territories

At the time of Confederation in 1867, the Dominion of Canada had comprised the provinces of Ontario, Québec, New Brunswick and Nova Scotia.[77] In drafting of the British North America Act 1867, the Fathers of Confederation had anticipated that the North-Western Territory and Rupert's Land would, at some point in the foreseeable future, join Confederation along with the rest of British North America.[78] They therefore made provision for this under section 146 by providing Parliament with the authority to admit the additional territories into Confederation.[79] Around a year after Confederation, Rupert's Land Act 1868 was enacted by the Imperial Parliament, giving the Hudson's Bay Company legal permission to transfer

[75] Yee (n 74); British North America Act 1867, ss 91-92.
[76] (1989) 2 SCR 49, 103.
[77] Conrad (n 28) 150; British North America Act 1867, s 5.
[78] Andrew McIntosh and Shirlee Anne Smith, 'Rupert's Land' (*The Canadian Encyclopadia*, 7 February 2006) <www.thecanadianencyclopedia.ca/en/article/ruperts-land> accessed 18 March 2021.
[79] Lewis H Thomas, 'Confederation and the West' [1968] Revista de Historia de América 33, 33; British North America Act 1867, s 146.

Rupert's Land to the Dominion.[80] Two years later, in 1870, the Deed of Surrender would be made by an order-in-council of Queen Victoria, officially transferring the North-Western Territory and Rupert's Land to Canada.[81] Initially, the territories were not considered provinces and consequently possessed no legislative rights under the section 92 of the British North America Act 1867.[82] Although they eventually acquired representation in Parliament under the British North America Act 1886, this was to be the extent of the territories' rights.[83]

Instead of being entitled to legislatures of their own, the territories were placed under the direct control of Parliament. According to section 4 of the British North America Act 1871, '[t]he Parliament of Canada may from time to time make provision for the administration, peace, order, and good government of any territory'.[84] Contrary to the implication, Parliament was under no legal obligation to legislate in a manner consistent with the objectives of 'peace, order, and good government'. Instead, Parliament was afforded absolute legislative discretion over the territories. This was confirmed by the Judicial Committee of the Privy Council in *Riel v The Queen*.[85]

Riel amounts to confirmation that the Parliament of Canada has indivisible legislative sovereignty over the territories and possessed the absolute authority to legislate on their behalf. Consequently, the sovereignty of the Parliament of Canada is at its strongest in the territories as it is not impeded by the principle of federalism in these regions. Canadian parliamentary sovereignty can therefore be understood as a region-dependent principle, as opposed to its British counterpart, which is universal.

---

[80] Rupert's Land Act 1868, ss 3-5.
[81] *Order of Her Majesty in Council admitting Rupert's Land and the North-Western Territory into the union, dated the 23rd day of June 1870* (Windsor, 23 June 1870).
[82] Katherine A Graham and Eli Yarhi, 'Territorial Government in Canada' (*The Canadian Encyclopaedia*, 7 February 2006) <www.thecanadianencyclopedia.ca/en/article/territorial-government> accessed 28 March 2021.
[83] British North America Act 1886.
[84] British North America Act 1871, s 4.
[85] [1885] 10 AC 675, 677-79.

Although under no legal obligation to do so, the Parliament of Canada has, under its delegatory powers, established devolved legislatures for the territories (not unlike those of Scotland, Wales and Northern Ireland).[86] In *R v Burah*, a case appealed from the British Raj, the House of Lords ruled that colonial legislatures possessed delegatory powers 'as large and of the same nature as those of the Imperial Parliament itself'.[87] The House of Lords would later affirm this in *Hodge v R* for both the Parliament and provincial legislatures of Canada, ruling that their authority to delegate legislative power is both 'plenary and ample'.[88] As a consequence of this relatively unrestricted delegative authority, the Canadian Parliament has devolved a significant legislative and executive authority to the territorial legislatures. Indeed, the foundational legislation of territorial legislatures contains lists of areas of legislative competence that closely resemble section 92 of the British North America Act 1867.[89]

It should be noted, however, that although both Parliament and the provincial legislatures possess the authority to delegate legislative authority, the power is not absolute. The Supreme Court of Canada has ruled in *Re Gray* that any 'abdication', 'abandonment' or 'surrender' of federal or provincial power would be invalid.[90] This is because such a move would amount to Parliament permanently destroying its own sovereignty. This is incompatible with the doctrine of parliamentary sovereignty and with the British North America Acts 1867, section 91 and 1871, section 4.

Like Westminster, therefore, the Parliament of Canada possesses a legally unbridled right to abolish or restrict any of its devolved legislatures, although exercising this right would not be without its political difficulties. Consequently, the devolved legislatures cannot be considered to be sovereign for the purposes of Canadian parliamentary sovereignty.

---

[86] Yukon Act, 2002; Northwest Territories Act, 2014; Nunavut Act, 1993.
[87] [1878] 3 AC 889, 904.
[88] [1883] 9 AC 117, 132.
[89] Yukon Act, 2002, ss 17-27; Northwest Territories Act, 2014, c 2, ss 9, 16; Nunavut Act, 1993, ss 12, 23.
[90] [1918] SCR 150, 157, 165, 171, 176.

Creating New Provinces

Although the British North America Act 1867 had provided the Dominion of Canada with the authority to admit the colonies of Prince Edward Island, British Columbia and Newfoundland into Confederation alongside the territories, it did not give the Canadian Parliament the authority to create new provinces.[91] Ignoring this, Parliament enected the Manitoba Act, 1870 in an attempt to create the Province of Manitoba out of Rupert's Land.[92] As there had been doubt as to the validity of this Act, the Parliament of the United Kingdom would later give retrospective effect to it by enacting the British North America Act 1871 in the following year.[93]

In order to avoid a similar situation in the future, the Parliament of Canada was given the express authority to establish new provinces in the territories under section 2 of the 1871 Act.[94] Furthermore, Parliament was given the authority to cede territorial land to an extant province under section 3 of the Act on the condition that it receives the province's consent before doing so. Exercising these powers, according to the British concept of parliamentary sovereignty, effectively amounts to the Parliament of Canada destroying its own sovereignty by permanently limiting its legislative authority over any land that it cedes.[95]

This is not so according to the Canadian concept of parliamentary sovereignty. In *Re Gray*, the Supreme Court of Canada held that by creating or expanding provinces, the Parliament of Canada is horizontally transferring its 'sovereignty' to the provincial legislatures, who are in and of themselves 'sovereign'.[96]

While the Parliament of Canada is entitled to pass laws for newly created provinces at the time of their creation,[97] it cannot do

[91] British North America Act 1867, s 146.
[92] Manitoba Act, 1870.
[93] British North America Act 1871, s 5; Frank L Bastedo, 'Amending the British North America Act' (1934) 4 Canadian Bar Rev 209, 216.
[94] Bastedo (n 93).
[95] Dicey (n 2) 26-30.
[96] [1918] SCR 150, 157, 165, 171, 176.
[97] British North America Act 1871, s 2.

this indefinitely. Section 6 of the British North America Act 1871 explicitly prohibits Parliament from legislating on a province's behalf after its establishment. Were it allowed to do so indefinitely, a new province would be incapable of ever becoming 'sovereign' and would function as a *de facto* territory for the purposes of Canadian parliamentary sovereignty. Consequently, section 6 has the effect of bringing newly established provinces within the scope of section 92 of the British North America Act 1867, making them sovereign over their own legislative affairs.

The Parliament of Canada went on to create Alberta and Saskatchewan in 1905.[98] These were the only two provinces created out of the territories using Parliament's powers under section 2 of the British North America Act 1871. The other provinces to join Canada post-Confederation were former colonies of British North America.[99]

The Balfour Declaration and the Statute of Westminster 1931

The Balfour Declaration of 1926[100] heralded a change to the sovereign status of the Parliament and provincial legislatures of Canada. Although of no legal effect in and of itself, it declared that the United Kingdom and the Dominions

> are autonomous Communities within the British Empire, equal in status, in no way subordinate one to another in any aspect of their domestic or external affairs, though united by a common allegiance to the Crown, and freely associated as members of the British Commonwealth of Nations. [101]

The impact that this statement had cannot be understated. Immediately, it had ended the subordinate status of the Dominions, making them *de facto* equals to the United Kingdom itself. It ended the role of the Governors General as the representatives of the British

[98] Saskatchewan Act, 1905; Alberta Act, 1905.
[99] Hogg (n 35) 2.12–.16.
[100] Not to be confused with the Balfour Declaration of 1917 which announced the British government's support for establishing a Jewish homeland in Palestine.
[101] Inter-Imperial Relations Committee, 'Report, Proceedings and Memoranda' (Imperial Conference 1926, London, November 1926).

government and granted the Dominions the authority to make treaties and declare war.[102] According to van Themaat, although the Balfour Declaration amounted to a declaration of a new constitutional convention, it was not in and of itself a legal instrument and was therefore of little to no legal effect, despite its political ramifications.[103] The declaration was effectively a political guarantee from the United Kingdom that it would no longer exercise political or legal control over its Dominions without their consent.

It was agreed at the Imperial Conference of 1930 that the Imperial Parliament would give legal effect to the Balfour Declaration and give the Dominions legislative independence.[104] This occurred the following year with the Statute of Westminster 1931. As the Ontario Court of Appeal explains,

> the Statute of Westminster was a significant development for Canadian sovereignty, in that it permitted Canada to pass laws that were inconsistent with certain British laws for the first time.[105]

Section 7 of the Statute of Westminster provided for the continued supremacy of the British North America Acts in Canada. However, section 2(1) of the Statute disapplied the Colonial Laws Validity Act 1865 and section 2(2) explicitly granted the provincial legislatures and the Parliament of Canada with the authority to legislate over *all* matters not directly addressed by the British North America Acts. Sections 3, 5 and 6 of the Statute gave Canada the express authority to legislate extra-territorially and over matters pertaining to the Courts of Admiralty and merchant shipping. Taken together, these provisions rendered the Dominion's legislative bodies collectively 'sovereign' over all matters save the ability to amend its own constitution.

---

[102] Mazer (n 44) 11.
[103] H ver Loren van Themaat, 'Legislative Supremacy in the Union of South Africa' (1954) 3 University of Western Australia L Rev 59, 62-63.
[104] Hogg (n 35) 3-5.
[105] *McAteer v Canada (Attorney General)* [2014] ONCA 578 [45].

Canada had near-absolute internal autonomy from Confederation[106] and the new law-making powers it had acquired under the Statute of Westminster were almost exclusively of an external nature. Therefore, the Statute of Westminster primarily expanded the scope of the Canadian Parliament's sovereignty, rather than that of the provincial legislatures.

Section 7 of the Statute of Westminster reserved the Imperial Parliament's power to amend the British North America Acts. Why was this? Other Dominions, such as Australia,[107] were given the power to amend their own constitutions upon their creation. Canada was not. This was on account of the Fathers of Confederation being satisfied to reserve this power to the Imperial Parliament. Yet why, even in 1931, was it not handed over to Canada? The reason was that the federal and provincial governments were unable to settle on an agreeable amendment procedure.[108] Consequently, Westminster would retain the power to amend the British North America Acts for another 51 years.

By strengthening the principle of parliamentary sovereignty in the Dominion of Canada, the Parliament of the United Kingdom was doing what was traditionally thought to be impossible: restricting its own sovereignty and binding its successors.[109] According to section 4 of the Statute of Westminster, Acts of the British Parliament are incapable of extending to any Dominion, unless such an act expressly declares 'that that Dominion has requested, and consented to, the enactment thereof.'[110] It seems unlikely that the British courts would ever have struck down primary legislation that violates this legal guarantee. In the case that they had, or if the UK had attempted to repeal the Statute of Westminster or unilaterally amend the British North America Acts, Mazer argues that 'Canada would [have] unilaterally declare[d] independence or at least [have made] immediate moves to patriate the constitution'.[111]

---

[106] Mazer (n 44) 11.
[107] For example, Commonwealth of Australia Constitution Act 1900, s 128.
[108] Forsey (n 29) 9-10.
[109] Dicey (n 2) 26-30.
[110] Statute of Westminster 1931, s 4.
[111] Mazer (n 44) 14.

The enactment of the Statute of Westminster in 1931, marked the zenith of the sovereignty of the Parliament of Canada and the provincial legislatures. The only permanent limitations on the principle of parliamentary sovereignty were being the British North America Acts, which were primarily concerned with federalism.

Recognising Quasi-Constitutional Legislation

The enactment of the Canadian Bill of Rights by the Parliament of Canada in 1960 would set in motion a series of events that would lead to the recognition of quasi-constitutional legislation by the courts, which would act as a 'soft' limitation on the sovereignty of the provincial legislatures and Parliament.[112]

According to the doctrine of implied repeal, a doctrine which Canada has inherited from the United Kingdom, no two statutes can be of simultaneous legal force if they are inconsistent or in conflict with one another. If such an inconsistency or conflict is found to exist between two statutes, then the earlier statute is considered to have been repealed to the extent that it conflicts with the latter statute upon the latter statute's enactment.[113]

This doctrine was challenged by the Canadian Bill of Rights. Section 2 of the Bill of Rights purports to apply to 'every law of Canada', requiring that they be construed so 'as not to abrogate, abridge or infringe or to authorize the abrogation, abridgment or infringement of any of the rights or freedoms' contained with the Bill.[114] Under the traditional understanding of the doctrine of implied repeal, the Canadian Bill of Rights should, as a seemingly ordinary Act of Parliament, have applied only to previous legislation. This would have been consistent with the traditional understanding of parliamentary sovereignty that Canada had inherited from the United Kingdom: that there was to be 'no marked or clear distinction between laws which are not fundamental or constitutional and laws which are

[112] Canadian Bill of Rights, 1960.
[113] Canada v Schmidt (1987) 1 SCR 500.
[114] *Canadian Bill of Rights*, 1960, s 2.

fundamental or constitutional', aside from the British North America Acts.[115]

However, the Supreme Court of Canada departed from the doctrine of implied repeal in *R v Drybones*, ruling that the Canadian Bill of Rights was binding on all federal legislation, past, present and future.[116] Somewhat surprisingly, the Supreme Court went one step further than requiring that future legislation be construed in a manner consistent with the Bill.[117] They ruled that legislation that could not be interpreted to give effect to the rights and freedoms under the Canadian Bill of Rights was to be considered void.[118] This later occurred in *Singh v Minister of Employment and Immigration*.[119] Although the Supreme Court's ruling did not entrench the Canadian Bill of Rights, it had the effect of making the Bill immune from implied repeal. In order to legislate notwithstanding the Canadian Bill of Rights Parliament must make an express declaration in a repugnant Act of its intention to do so. This has only been done once, with the Public Order (Temporary Measures) Act during the 1970 'October Crisis' involving the Front de libération du Québec.[120]

Eventually, Laskin J coined the term 'quasi-constitutional legislation' in *Hogan v R* to refer to legislation, such as the Canadian Bill of Rights, which cannot be impliedly repealed.[121] As its name implies, quasi-constitutional legislation cannot override the Constitution of Canada as doing so would amount to a constitutional amendment. This cannot be done by ordinary legislation. Quasi-constitutional legislation cannot hop the federal-provincial divide in violation of federalism. Nevertheless, the recognition of quasi-constitutional legislation's semi-binding effect would amount to an alteration to Canada's rule of recognition and would have major implications for parliamentary sovereignty. By recognising the existence of quasi-constitutional legislation, the Supreme Court of Canada recognised that Parliament and the legislatures were capable

[115] Dicey (n 2) 37.
[116] [1970] SCR 282.
[117] Hogg (n 35) 35-4.
[118] *R v Drybones* (n 116).
[119] (1985) 1 SCR 177.
[120] Public Order (Temporary Measures) Act, SC 1970, s 12; Hogg (n 35) 35-4.
[121] (1975) 2 SCR 574, 597-58 (Laskin J).

of partially binding themselves and their successors. This had hitherto been thought to be impossible without formal constitutional amendment.

Although repugnant with the orthodox British concept of parliamentary sovereignty, the British courts would later come to recognise 'constitutional statutes' (for example, in *Thoburn v Sunderland City Council*, the Human Rights Act 1998). This imposed a similar 'soft' limitation on parliamentary sovereignty in the UK.[122] Although Westminster cannot repeal such statutes by implication, the UK courts never went as far as the Canadian courts did by ruling that they are capable of rendering subsequent statutes inoperative.

The Implications of Patriating the Canadian Constitution

Obtaining Full Independence from the United Kingdom

Although Canada had acquired near-absolute autonomy from the United Kingdom under the Statute of Westminster in 1931, it was not until the enactment of the Canada Act 1982 by the Parliament of the United Kingdom that Canada finally achieved complete independence.[123] The Canada Act was the product of a years-long political process known as 'patriation' in which control of the Constitution of Canada was finally handed from Westminster to the Parliament and provincial legislatures of Canada.[124]

Under section 2 of the Canada Act 1982, Westminster terminated all legislative authority over Canada. It declares: '[n]o Act of the Parliament of the United Kingdom passed after the Constitution Act, 1982 comes into force shall extend to Canada as part of its law.' Although this appears to grant Canada unconditional independence, it is also apparent that Westminster was once again doing what had traditionally been considered impossible: destroying its own sovereignty and binding its successors.[125] While it is understood that

[122] [2002] EWHC 195 (Admin).
[123] Barbara Billingsley, 'Evolution, Not Revolution: Canada's Constitutional History and the Constitution Act, 1867' (2013) 37 LawNow 8, 10-11.
[124] Forsey (n 29) 10.
[125] Dicey (n 2) 21.

British parliamentary sovereignty is compatible with 'soft' limitations, as was discussed in chapter one, no 'hard' limitations such as this have ever been recognised by the UK courts. This raises an important question: are the Parliament and provincial legislatures of Canada still legally subject to the authority of Westminster?

According to Hogg, by eliminating its own role in Canada's law-making process Westminster has 'made Canada's legal system technically self-sufficient – perfectly independent of the United Kingdom.'[126] If Westminster were to try and legislate for Canada post-patriation, Hogg asserts that the Canadian courts would consider such laws invalid and of no force or effect.[127] Hogg's argument is supported by the South African case of *Ndlwana v Hofmeyr*, in which the Appellate Division, South Africa's highest court, ruled that the legal system had ceased to recognise the authority of the Imperial Parliament following the Statute of Westminster.[128] The Appellate Division's reasoning was that '[f]reedom, once conferred, cannot be revoked.'[129]

Canada achieved full formal independence from the United Kingdom upon the enactment of the Canada Act 1982. The Act rendered Parliament and provincial legislatures of Canada entirely sovereign over matters falling within their respective scopes of authority, free of any constraints aside from the Constitution of Canada itself. This put an end to the notion that Canadian autonomy is 'sovereignty.' It also marked a final change to Canada's rule of recognition, which can presently be stated as follows: the Parliament and provincial legislatures of Canada shall be sovereign to the extent that they legislate consistently with the written Constitution of Canada.

Amendment Procedure

Although the Parliament of Canada had initially been given limited powers to amend the Canadian constitution under the provisions of

[126] Hogg (n 35) 3.14-.15.
[127] ibid.
[128] [1937] AD 229.
[129] ibid 237.

the British North America (No 2) Act 1949, it was not until patriation that Canada would formally be in full control of its procedure for constitutional amendment.[130] The Constitution Act, 1982 sets out five different procedures of constitutional amendment.[131] Each deals with the amendment of different aspects of the constitution. Unlike the British constitution, which is neither codified nor supreme, the Canadian constitution cannot be amended using by ordinary legislation Acts. The Canadian Parliament and provincial legislatures must instead follow the amending procedures set out in the Constitution Act, 1982. Attempting to amend the constitution outwith these procedures would be held to be *ultra vires* and of no legal effect.[132]

Section 44 of the 1982 Act provides the Parliament of Canada with unilateral authority to amend any constitutional provisions relating to the federal executive, Senate and House of Commons.[133] Ssection 45 provides each of the provincial legislatures with the authority to amend their own province's constitution unilaterally.[134] These two provisions render the Parliament of Canada and the provincial legislatures *inwardly* sovereign; sovereign over all constitutional matters only applicable to themselves. This enables each legislature to change their its legislative procedures, similar to the UK Parliament under the Parliament Acts of 1911 and 1949.

Unlike the constitutions of countries like Switzerland and Germany), the Constitution of Canada does not contain an 'eternity clause'.[135] Unlike a supremacy clause, which provides for a constitution's supremacy over all other laws, an eternity clause prohibits certain amendments from ever being made to a constitution under any circumstances and has the effect of permanently entrenching a provision into a constitution.[136] Were Canada to have such a clause in its constitution, it would be unable to amend any

[130] British North America (No 2) Act 1949, s 1.
[131] Constitution Act, 1982, pt 5.
[132] ibid s 52(1).
[133] ibid s 44.
[134] ibid s 45.
[135] Ondřej Preuss, 'The Eternity Clause as a Smart Instrument – Lessons from the Czech Case Law' (2016) 57 Hungarian Journal of Legal Studies 289.
[136] ibid.

'eternal' provisions, even using the unanimous consent procedure, which requires the consent of Parliament and all ten provincial legislatures.[137]

It is therefore possible for all ten provincial legislatures and the Parliament of Canada acting as one collective unit to pass any constitutional amendment, and by extension, any law.[138] In other words, the Parliament of Canada and the provincial legislatures are *collectively sovereign* in the sense that no hard constitutional limitation whatsoever is placed on them when they act as one. Just as the British Parliament exercised its *unilateral* sovereignty by repealing the European Communities Act 1972,[139] the Canadian Parliament and the provincial legislatures are *collectively* sovereign in their power to amend or repeal any provision of the Constitution of Canada that limits their sovereignty. Theoretically, if Parliament and the provincial legislatures so wished, they could abolish Canada's democratic system of government, judicial system or the constitution itself.

The Impact of the Canadian Charter of Rights and Freedoms

The Charter and Parliamentary Sovereignty

The greatest impact of patriation on Canadian parliamentary sovereignty was not the patriation of the constitution itself, but the Canadian Charter of Rights and Freedoms. This was enacted as part of the Constitution Act, 1982.[140] Unlike the Canadian Bill of Rights or the United Kingdom's Human Rights Act 1998, the Canadian Charter of Rights and Freedoms is not an Act of Parliament. Instead, it is part of Canada's written constitution. Consequently, the Charter is afforded supremacy over the uncodified portion of the Canadian constitution, in addition to Acts of Parliament and the provincial legislatures, by virtue of section 52(1) of the Constitution Act, 1982.[141]

[137] Constitution Act, 1982, s 41.
[138] Constitution Act, 1982, pt 5.
[139] European Union (Withdrawal) Act 2018, s 1.
[140] Constitution Act, 1982, pt 1.
[141] Hogg (n 35) chs 36, 37, 40; Forcese and others (n 1) ch 11; Bird (n 36) 768.

However, the adoption of the Charter send Canada down an American style jurisprudential path based almost entirely on the codified human rights portion of its constitution.[142] It did not abolish the principle of parliamentary sovereignty entirely. Instead, as Kazmierski points out, Canadian judges have continued to adopt

> a "British" approach that recognises the legitimacy of unwritten principles but favours the principle of parliamentary sovereignty above other principles' [in their interpretation of the constitution.][143]

This is supported by the comments of Strayer J in *Singh v Canada (Attorney General)*, who states that:

> Both before and after 1982 our system was and is one of parliamentary sovereignty exercisable within the limits of a written constitution. These were solely quantitative limits on the exercise of legislative power prior to 1982. It is true that the adoption of the Charter in 1982 added a multitude of qualitative limitations on the exercise of power, but it is difficult to ascertain any change in the principle that the Constitution of Canada was and is supreme over ordinary laws. As a result one is driven as before 1982 to looking at the specific requirements of the Constitution to determine whether in a given case Parliament has infringed a constitutional limit (express or implied) on its power.[144]

However, it cannot be denied that the Charter materially expanded the role of Canada's written constitution at the expense of its unwritten constitution. The Supreme Court of Canada observed in in *Reference Re Secession of Quebec* that the adoption of the Charter significantly transformed 'the Canadian system of government… from a system of Parliamentary supremacy to one of constitutional supremacy.'[145]

---

[142] Kazmierski (n 51) 245.
[143] ibid; *Bacon v Saskatchewan Crop Insurance Corp* (1990) 180 SaskR 20, [29].
[144] (1999) 4 FC 583, [16] (Stayer J).
[145] (1998) 2 SCR 217, [72].

Consequently, the principle of parliamentary sovereignty plays a much more minor role in the Canadian constitutional order following the patriation of the constitution. The scope of the legislative authority of both Parliament and the provincial legislatures has been significantly reduced by the rights and freedoms enumerated in the Charter.[146]

The Notwithstanding Clause: Preserving Parliamentary Sovereignty?

It is commonly purported that section 33 of the Charter of Rights and Freedoms 'preserves' parliamentary sovereignty.[147] The basis of this argument is that section 33 allows Parliament and the provincial legislatures to enact legislation that would otherwise be void for violating the Charter.[148] Section 33 provides Parliament and the provincial legislatures with the authority to legislate notwithstanding sections 2 and 7 through 15 of the Charter by making an express declaration under this section in an Act.[149] It is commonly known as the 'notwithstanding clause'.

Section 33 of the Charter of Rights and Freedoms is similar in a sense to section 2 of the Canadian Bill of Rights, which also requires an express declaration to be made by the Canadian Parliament or provincial legislatures in order to legislate in contravention of the Bill. However, the Charter's notwithstanding clause goes further. Section 33(3) imposes a mandatory 5-year sunset clause on all declarations of incompatibility. Section 33(4) requires Parliament to re-enact a declaration to reset the 5-year sunset clause. Because of this mandatory sunset clause, section 33 cannot be considered the preservation of parliamentary sovereignty nor an extension of the principle.

Canadian parliamentary sovereignty consists of the ability of Parliament and the legislatures to make or unmake any law within

---

[146] For a recent case involving the principle, see *Reference re Pan-Canadian Securities Regulation* [2018] SCC 48.
[147] Goldsworthy (n 43) ch 8.
[148] The Supreme Court of Canada affirmed that s 33 allows for the enactment of repugnant legislation in *Ford v Quebec (Attorney General)* (1988) 2 SCR 712.
[149] Constitution Act, 1982, Sch B, Pt 1, para 33.

their respective scopes of authority. Ultimately, 'no person or body [is] recognised… as having a right to override or set aside… the legislation of Parliament'.[150] Canadian parliamentary sovereignty entails a legislative body having ultimately unconditional law-making authority over its affairs.

The sunset clause imposed by section 33(3) of the Charter means that Parliament and the provincial legislatures are not sovereign when legislating in certain areas. Instead, they are conditionally legislating and are always subject to the supremacy of the constitution, which is directly applicable. As Vanessa MacDonnell argues,

> characterizing the notwithstanding clause as preserving sovereignty is inconsistent with the only descriptively and normatively plausible variant of parliamentary sovereignty in Canadian constitutional law: a limited sovereignty conditioned by conditional rights.[151]

The notwithstanding clause neither preserves Canadian parliamentary sovereignty nor the rights and freedoms enumerated in the Charter. It should be understood to be a hybrid expression of Canada's uncodified and codified constitutional traditions that conditions rights and freedoms as opposed to 'saving' or 'preserving' parliamentary sovereignty.

Conclusion

The doctrine of parliamentary sovereignty was successfully adopted into the Canadian legal order at the time of Canadian Confederation in 1867. This was seemingly an impossibility on the basis of the orthodox British concept. Adjustments were necessary in order to accommodate for Canada's federal system of government, partially codified constitution and subordinate status to the UK. However, Canadian parliamentary sovereignty nevertheless came to thrive in its own right; defying all British accounts of the principle.

---

[150] Dicey (n 2) 3-4.

[151] Vanessa MacDonnell, 'The New Parliamentary Sovereignty' (2016) 21 Review of Constitutional Studies 13, 25.

Between 1867 and 1931, Canadian parliamentary sovereignty continued to expand with Canada's acquisition of the territories. The Canadian Parliament and provincial legislatures achieved near-absolute legal independence following the enactment of the Statute of Westminster in 1931 by the UK Parliament.

The inviolability of Canadian parliamentary sovereignty was dented by the Supreme Court of Canada's recognition of quasi-constitutional legislation in 1970. Nevertheless, the uncodified doctrine remained second to the principle of federalism in Canada's constitutional order until 1982.

The patriation of the Canadian constitution in 1982 would free Parliament and provincial legislatures from their *de jure* subordination to the British Parliament. However, the concurrent introduction of the Canadian Charter of Rights and Freedoms would relegate the principle to the edge of Canada's constitutional order, which would come to be dominated by the Charter. Although sidelined, parliamentary sovereignty remains the key uncodified principle of the Canadian constitution.

In the 250 years since Confederation, Canadian parliamentary sovereignty has evolved significantly from its British doctrinal roots. Canadian parliamentary sovereignty has come to be seen as a unique constitutional principle in and of itself. However, the time of Canadian parliamentary sovereignty has passed. Only its legacy remains.

Convexity Ltd v Phoenixfin Pte Ltd: *When Application of the Penalty Rule Clashed with Natural Justice*

ILIAS KAZEEM*

Abstract

This is a review of the Convexity *case, which shows how the application of the rule against penalty, under English law, must respect the principles of natural justice. Although the principles entail independence of an adjudicator and fair hearing of the parties, the case amplifies the relevance of the fair hearing limb. It presents an example of how international arbitration interacts with institutional rules in the governing law. In this regard, the case examines the mandatory requirement of fair hearing as a limit to the flexibility of arbitration. It also raises the question of the impact of lack of pleading of the penalty rule. In this case review, the facts of the* Convexity *case are presented in summary and the final award explained. The court judgment that nullified the award is also stated, before showing the analysis of the clash between the penalty rule and the requirement of fair hearing.*

Keywords: penalty rule, natural justice, international arbitration

Introduction

This case review considers the judgment of the High Court of Singapore (Court) delivered on 19 April 2021 in *Convexity Ltd v Phoenixfin Pte Ltd (Convexity).*[1] In this case, the Court set aside an arbitral award on the sole basis that the application of the penalty rule by the arbitral tribunal (Tribunal) breached the claimant's right to natural justice. The award had been rendered in Singapore, based on the expedited procedure of the rules of the Singapore International

* LLB, University of Ilorin; LLM, University of Aberdeen. The author is a barrister and solicitor in Nigeria. He is currently a PhD student and teaching fellow at the University of Aberdeen.

[1] 2021 SGHC 88.

Arbitration Centre (SIAC).[2] While other issues arose, the case review focuses on the penalty rule in connection with natural justice, which ultimately formed the reason for the decision.

Relevant Facts[3]

The parties entered a Service Agreement with an initial term of 24 months and a monthly fee of $200,000 payable to the claimant, for the services rendered under the contract. The contract was governed by English law. It provides that in the event of termination prior to the expiration of the initial term, the first respondent[4] (respondent) is obligated to pay a 'make-whole amount' to the claimant. This is any amount that will make the total payments received by the claimant to be the worth of the total monthly fees. Upon termination by the respondent, the claimant claimed the make whole amount of $2.8 Million and 5% interest.

At arbitration, both parties concluded pleadings without raising the issue of penalty rule under English law (penalty issue). The respondent subsequently requested to call an expert witness to testify on the penalty issue. The claimant opposed the application and suggested that any issue of English law 'arising in this arbitration' should be addressed in the closing submissions. The respondent applied at this stage to amend its pleadings, in order to plead the penalty issue, but based on the opposition by the claimant, the tribunal refused the application.

Consequently, the hearing was concluded, without the penalty issue. The tribunal nevertheless raised the penalty issue at closing submission stage several times, while the claimant consistently objected. The crux of the claimant's objection was that the issue was introduced by the respondent late and it was not pleaded as part of the scope of the dispute. The claimant contended that the penalty issue did not arise in the arbitration.

---

[2] SIAC Arbitration Rules, 2016 Art 5

[3] *Convexity* (n 1) at paras 1-26.

[4] The second and third respondents are parties to the case only as guarantors and indemnitors and the focus is on the first respondent.

The Final Award

The tribunal dismissed the claimant's claim, on the sole ground that the contractual provisions amounted to an unenforceable penalty under English law.[5] The tribunal found that the claimant had *agreed* that the penalty issue should be addressed at the closing submission stage of the arbitration.[6] It observed that the issue of pleading is not sacrosanct, since arbitration does not require the formalities that are applicable to litigation.[7] Thus, the tribunal applied the penalty rule. It is also noted that the award was rendered 3 months after the deadline for rendering an award under the expedited procedure.[8]

The Court Ruling Setting Aside the Award

The claimant argued that the late introduction of the penalty issue and the failure of the tribunal to allow the claimant's objection amounted to breach of natural justice. The Court started its analysis on the basis that the penalty rule is a matter of mixed law and fact, to be proved.[9] As a result, the Court held that it is required that the penalty issue must be pleaded.[10] In this case, the Court had no difficulty in holding that the penalty issue was not pleaded by the respondent, since the failure to plead was not in contention.[11] The Court added that the respondent had conceded the point that the penalty issue requires pleading, when it applied to amend its pleadings, in order to plead same.[12]

Furthermore, the Court found that the claimant did not agree to the late introduction of the penalty issue without pleadings, and the tribunal committed an error in failing to consider the persistent

[5] *Convexity* (n 1) at para 27.
[6] ibid at para 37.
[7] ibid.
[8] ibid at para 112.
[9] ibid at para 33. See also *Lombard North Central plc v European Skyjets Ltd (in liquidation) and another* (2020) EWHC 679 (QB) and *Banner Investment Pte Ltd v Hoe Seng Metal Fabrication & Engineers (S) Pte Ltd* (1996) 3 SLR(R) 244.
[10] *Convexity* (n 1).
[11] ibid at para 32.
[12] ibid at para 34.

objection of the claimant.[13] The Court established that the claimant's agreement was limited to matters of English law 'arising in this arbitration',[14] and not the penalty issue, which did not arise in the arbitration.[15]

Therefore, the Court held that the tribunal breached the principle of natural justice, as it failed to give due consideration to the claimant's objection or allow the claimant to fully argue it.[16] The Court considered that the tribunal was distracted by the penalty issue from the agreed expedited procedure.[17] The procedure spelt out the stages of the arbitration,[18] but the tribunal entertained protracted proceedings and it failed to meet the deadline for rendering the award.[19]

In defence of the award, the Court considered the respondent's arguments that the tribunal had a power to raise an issue of its own motion.[20] The respondent also argued that the tribunal has an obligation to protect public policy.[21] However, the Court found that the penalty issue being a question of mixed law and fact, the tribunal was not entitled to raise it without pleading.[22] While the Court agreed that the penalty issue is a matter of public policy, it found that the issue must be properly placed before an adjudicator.[23] The Court also held that the tribunal breached equality of treatment of the parties and it failed to provide an opportunity for the claimant to present its case.[24]

---

[13] ibid at para 57.
[14] ibid at para 41.
[15] ibid at paras 41, 43 and 54.
[16] ibid at paras 86-88. See also *Front Row Investment Holdings (Singapore) Pte Ltd v Daimler South East Asia Pte Ltd* 2010 SGHC 80.
[17] ibid at paras 112-116.
[18] ibid.
[19] ibid.
[20] ibid at paras 90-91.
[21] ibid. See also *PT Prima International Development v Kempinski Hotels SA and other appeals* (2012) 4 SLR 98.
[22] ibid at para 92. See also *Cavendish Square Holding BV v El Makdessi and another appeal* (2016) 2 All ER 519 at 7 and 9; *Denka Advantech Pte Ltd and another v Seraya Energy Pte Ltd and another and other appeals* 2020 SGCA 119 at 175.
[23] ibid. See also *Banner Investment case* (n 9).
[24] ibid at para 93. See also *China Machine New Energy Corp v Jaguar Guatamela LLC and another* (2020) 1 SLR 695.

The Court, therefore, set aside the award and requested for further hearing on the consequential order to be made in the case.[25]

Penalty Rule versus Natural Justice in Arbitration

It is important to consider the penalty rule and the principles of natural justice, in order to examine the significance of the interaction between the duo in this case.

The rule against penalty is a rule under English law which prohibits the imposition of a penalty by the parties for contractual breach.[26] English law permits liquidated damages but it prohibits penalty *in terrore*.[27] The contention has always been how to distinguish between liquidated damages and penalty. The test which the English court previously laid down in *Dunlop* case[28] is that an amount which is beyond the genuine pre-estimate of the damage to be suffered by the injured party constitutes a penalty.[29] In the recent case of *ParkingEye Limited v Beavis*,[30] the Supreme Court of the United Kingdom overruled the Dunlop test and provided a new test which first classifies contractual provisions into primary and secondary obligations. The apex Court held that primary obligations do not amount to penalty but a secondary obligation amounts to a penalty if it 'imposes a detriment on the contract-breaker out of all proportion to any legitimate interest of the innocent party in the enforcement of the primary obligation.'[31] It is noted that the circumstances of each

[25] ibid at paras 122 and 123.
[26] Andrew Ham, 'The rule against penalties in contract: An economic perspective' (1990) 17 Melbourne University Law Review 649 at 650-651. See also Judith Aldersey-Williams, Philip Spencer Ashley, Emma Riddle and others, 'Default Clauses in Joint Operating Agreements: Recent Guidance from English Courts' (2016) 2 International Energy Law Review 36.
[27] *Dunlop Pneumatic Tyre Co. Ltd. v New Garage &Motor Co. Ltd* 1915 AC 79. See also Andrew (n 26).
[28] ibid.
[29] *Campbell Discount Co Ltd v Bridge* 1962 AC 600 and *Alfred Mcalpine Capital projects Limited v Tilebox Limited* (2005) EWHC 281 (TCC).
[30] 2015 UKSC 67, combined with *El- Makdessi v Cavendish Square Holdings BV* 2015 UKSC 67.
[31] ibid at para 32. See also Damian Crosse, 'Practical implications of penalty clauses in English law' (18 July 2018) <www.pinsentmasons.com/out-

case are considered in determining this test.[32] This underscores the importance of pleadings in articulating the relevant facts and circumstances.

On the other hand, the notion of natural justice originated from the English common law and it comprises two major principles that focus on the independence of an adjudicator and the fair hearing of the parties.[33] The fair hearing part dictates that parties must be provided an opportunity to be heard before a decision can be made.[34] Accordingly, natural justice refers to specific matters of procedural rights and its contours can only be discerned from the context and circumstances of each case.[35]

In arbitration, natural justice may not be expressly mentioned in the applicable rules.[36] However, the provisions for equal treatment of the parties and the opportunity to present one's case entail the fair hearing limb of natural justice.[37] Thus, natural justice principles have become indispensable requirements for arbitration as a form of adjudication.[38] An award can be set aside for breach of natural justice, where the party relying on the breach can clearly establish same and show the prejudice suffered by him.[39] This is distinguished from

---

law/guides/practical-implications-penalty-clauses-english-law> accessed 25 April 2021.

[32] Jonathan Morgan, 'Penalty clause doctrine: unlovable but untouchable' (2016) 75 The Cambridge Law Journal 11 at 23. See also Scott Crichton Styles, 'Joint Operating Agreements', in Greg Gordon, John Paterson and Emre Usenmez (eds), *UK Oil and Gas Law: Current Practice and Emerging Trends* (3rd edn, Vol II, Edinburgh University Press 2018) 59-60.

[33] Khushboo Hashu Shahdadpuri, 'The Natural Justice Fallibility in Singapore Arbitration Proceedings' (2014) 26(2) Singapore Academy of Law Journal 562.

[34] ibid at 563 and 564. See also *Gas and Fuel Corporations of Victoria v Wood Hall Ltd & Leonard Pipeline Contractors Ltd* (1978) VR 385 at 396.

[35] Stephen R Tully, 'Challenging Awards Before National Courts for a Denial of Natural Justice: Lessons from Australia' (2016) 32 Arbitration International 659 at 661.

[36] Khushboo (n 33) at 565 and 566.

[37] ibid.

[38] Austin I Pulle, 'Securing Natural Justice in Arbitration Proceedings (2012) 20(1) Asia Pacific Law Review 63 at 65.

[39] Khushboo (n 33) at 567 and 568. See also Koo Zhi Xuan and Joshua Lim Yong En, 'The Intricacies Involved in the Pursuit of Natural Justice in Arbitration' (2013) 25(2) Singapore Academy of Law Journal 595.

instances where a party merely seeks to use technical objections to defeat arbitration, without cogent grounds.[40] Consequently, while the formal requirements of court rules do not apply to arbitration, the process of arbitration is required to comply with the principles of natural justice.[41] The requirement is to be satisfied based on the nature of the dispute, the expectations of the parties and the character of the hearing.[42]

Conclusion

The *Convexity case* demonstrates that a tribunal invited to apply the penalty rule, as a matter of the governing law, needs to balance its application, with the requirements of natural justice. This is particularly the case where, as in the *Convexity case,* the facts to establish the test of the penalty rule have not been pleaded. The point that arbitration is less formal in comparison to court proceedings is not a justification for breach of natural justice. Tribunals must strive to balance flexibility, as a feature of arbitration, with the requirements of natural justice as fundamental to the process of any adjudication.

[40] ibid.
[41] Samantha Hepburn, 'Natural Justice and Commercial Arbitration (1993) 21(1) Australia Business Law Review 43, 47 and 48.
[42] ibid.

GUIDELINES FOR CONTRIBUTORS

The Aberdeen Student Law Review is published annually. The Editorial Board welcome submissions from students and recent alumni of the University of Aberdeen School of Law. A call for submissions containing further details will be issued in due course through School of Law internal communications, as well as on our LinkedIn and Twitter pages.

ASLR Contact Details

Aberdeen Student Law Review
c/o Law School Office
Room C41
School of Law
University of Aberdeen
AB24 3UB
United Kingdom

Web: www.abdn.ac.uk/law/student-activities/aberdeen-student-law-review-95.php

Email: aslr@abdn.ac.uk

Twitter: @AberdeenSLR

LinkedIn: Aberdeen Student Law Review