CAMBRIDGE LAW REVIEW

VOLUME IV
ISSUE I
SPRING 2019

Editor-In-Chief
Jared Kang

Proudly Supported By
Cambridge University Law Society

CAMBRIDGE LAW REVIEW

VOLUME IV
ISSUE I
SPRING 2019

Editor-In-Chief
Jared Kang

Proudly Supported By
Cambridge University Law Society

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

The Cambridge Law Review is now in its fourth year. It has been, and will continue to be a busy year for the journal—which will now publish two issues annually; in spring and autumn. Work has also been put towards developing a new website for the journal as well as its subsidiary undergraduate journal, De Lege Ferenda: www.cambridgelawreview.com.

The journal continued to benefit from a team of disciplined and dedicated editors and the guidance of our Honorary Board. I owe my thanks to each of them, and especially to the Vice Editors-in-Chief, Neli Frost and Rabin Kok; and this issue's tireless and exceptional Managing Editor, Hope Williams. Without them, this issue would not have been possible. We also welcomed a small number of international editors this year—students from universities in common law jurisdictions around the world. Many of submissions received in the few years since the journal was started have been from authors outside England and Wales, focused on domestic issues within their respective countries, many of which we can learn from generally. These international editors provided important insight and jurisdiction-specific knowledge where needed. They will continue to play a vital role in the issues and volumes to come, helping to ensure the highest standards of quality.

I would also like to thank the Cambridge University Law Society for their continued support, especially President for the 2018–2019 term, Gabriel Wang, for establishing long-term support that ensures the journal's independence and quality; the Editor-in-Chief for Volume III, Yen Jean Wee, for her invaluable and continued guidance; and Craig Slade of Crucible Creative for his excellent work on the publication.

With that said, I am pleased to present the Spring Issue of Volume IV, comprising legal scholarship in a variety of disciplines. From the Uttarakhand High Court's step towards ecocentrism in granting legal personality to the Ganga and Narmada rivers and their ecosystems; a rights-based analysis of the problems with Northern Ireland's restrictive abortion laws; an inquiry into the shortcomings of Hong Kong's legislative framework in manging the meaningful participation of accused persons suffering from mental, intellectual, or cognitive disabilities; to a paper which considers the recognition of 'irretrievable breakdown' as a ground for Christian divorce in Pakistan.

These papers may spark interest or inspiration for readers to go down similar, further, or comparative lines of inquiry, or they may not. However, their quality of scholarship and distance from the traditional subjects taught in law schools around the world serves actively as a reminder to students and young lawyers to broaden their intellectual horizons. Much can be gained from greater participation in areas where law, society, and environment intersect; areas in which there are no clear doctrinal answers, no elegant and theoretically satisfying solutions, and where the only way forward is through discourse, debate, and will. We are proud to publish such thoughtful pieces to this end.

I look forward to presenting the autumn issue later in the year.

Jared Kang
April 2019

Cambridge Law Review

Table of Contents

The Fair Trial Rights of Accused Persons Found 'Unfit to Plead and Stand Trial' in the Hong Kong Special Administrative Region

Andra le Roux-Kemp*

Abstract

It is a foundational principle of modern criminal justice systems that accused persons must possess the necessary faculties to effectively and meaningfully participate in criminal proceedings. In the laws of England and Wales, formal statutory recognition of this fair trial right first appeared in 1800 and has since remained an ongoing legislative project keeping abreast with contemporary understandings and awareness of mental health and cognition and reflecting interminable efforts to develop procedures that embolden fairness and justice. In this article, the legislative framework for "fitness to plead and stand trial" in the Hong Kong Special Administrative Region, is critically analysed with reference to the ongoing law reform and development in England and Wales. While the primary aim of this article is to critically evaluate whether and to what extent the relevant legislative framework of Hong Kong fosters or impedes the fair trial rights of accused persons suffering from a mental, intellectual, or other cognitive impairment, the discussion and analysis will also provide an opportunity for meaningful reflection on the evolution of fair trial rights for vulnerable accused in light of the adoption of the United Nations Convention on the Rights of Persons with Disabilities. Thus, in addition to revealing the current shortcomings of the Hong Kong legislative scheme, this article prompts renewed attention on the importance of *fairness* and *equality* in criminal proceedings involving vulnerable accused.

* Assistant Professor at the School of Law, City University of Hong Kong. BA LLB (Stell) CML (UNISA) LLD (Stell) BMus (UNISA) Hons BMus (UNISA).

I. Introduction

The right of an accused person to a fair trial entails, as a basic minimum, that such a person possesses of the necessary faculties to understand the charge against him or her, to plead thereto, and to participate in a meaningful and effective manner in any criminal proceeding that may ensue. To determine whether an accused person is indeed fit to plead and stand trial has, therefore, become an important part of contemporary criminal procedure. This, however, was not always the case. Fair trial rights, as such, were not always of primary concern and impetus in proceedings aimed at establishing whether an accused person is fit to plead and stand trial. At first, the primary concern was rooted in the possibility of the Crown to seize the property of an accused in the event of a conviction, and in the necessity of being able to distinguish "whether an accused was mute by malice or mute by visitation of God".[1] Both these concerns were aimed at serving the interests of the Crown in terms of execution of punishment; whether it be to punish upon conviction or for contempt of court. It was not until the eighteenth century that the focus shifted to the rights of the accused.

In 1790, Lord Chief Justice Keynon stated as follows: "No man shall be called upon to make his defence at a time when his mind is in that situation as not to appear capable of so doing."[2] Shortly after this pronouncement the first statutory provision for determining whether an accused person is fit to plead and stand trial was enacted in the laws of England and Wales under section 2 of the Criminal Lunatics Act 1800:

If any person indicted for any offence shall be insane, and shall upon arraignment, be found so to be by a jury lawfully impanelled for that purpose, so that such a person cannot be tried upon such indictment, or if upon the trial of any such person so indicted such person shall appear to the jury charged with such indictment to be insane, it shall be lawful for the court before whom any such person shall be brought to be arraigned or tried as aforesaid to direct such finding to be recorded, and thereupon to order such person to be kept in strict custody until his Majesty's please shall be known.

De Souza explains that this law was initially applied in several nineteenth-century trials involving "deaf-mutes" and later also in cases involving accused persons suffering from various forms of mental disorder.[3] Revisions to this first legislative

1 Law Commission, *Unfitness to Plead* (Law Com CP 197, 2010) [2.3].

2 ibid [2.5]; Nigel Walker, *Crime and Insanity in England and Wales: The Historical Perspective Vol. 1* (Edinburg: Edinburgh University Press 1968) p. 222.

3 Dominic S.M. De Souza, 'The concept of unfitness to plead' (Sep 2007) 9:3 The British Journal of Forensic Practice 7, 8.

enactment followed by way of the Criminal Procedure (Insanity) Act 1964,[4] the Criminal Procedure (Insanity and Unfitness to Plead) Act 1991,[5] and the Domestic Violence, Crime and Victims Act 2004,[6] with each revision the relevant procedures were further streamlined and aligned with the accumulation of knowledge and awareness of mental health and cognition, as well as the continuing expansion of fair trial rights in criminal procedure. This legislative project for fitness to plead and stand trial in the laws of England and Wales is still ongoing. In 2016, the Law Commission published an extensive report detailing recommendations for further legislative amendments,[7] together with the draft text of a Criminal Procedure (Lack of Capacity) Bill.[8] In addition to further procedural fine-tuning, this report and the draft bill also address, for the first time, the formulation of a new legal test for determining whether an accused person is fit to plead and stand trial.[9]

Similar legal development can also be noted in other parts of the Commonwealth. For example, in 2014, the Australian Law Reform Commission published a comprehensive inquiry into the laws and legal frameworks that directly or indirectly impact on the recognition of people with disabilities, including mental disabilities, and specifically how such persons can exercise their legal capacity on an equal basis with others. It was recommended in this report that the concept of fitness to stand trial be reformulated, "to focus on whether, and to what extent, a person can be supported to play their role in the justice system, rather than on whether they have the capacity to play such a role at all".[10] In South Africa, in turn, the Constitutional Court declared unconstitutional the default position provided for under the Criminal Procedure Act 51 of 1977, whereby an accused person having been declared unfit to plead and stand trial faced detention in a prison

4 For example, section 4 of this Act introduced a new procedure for determining whether an accused person is fit to plead and stand trial and also provided for a right of appeal against such a finding. The amendments in this Act was prompted by the 1963 report by the Criminal Law Revision Committee.

5 The amendments in this Act was prompted by the Report of the Committee on Mentally Abnormal Offenders (Cmnd 6244, 1975) (generally referred to as The Butler Report) and included, inter alia, for a mandatory hearing of the facts of the case once an accused has been found to be unfit to plead (section 4A).

6 Section 22(1) to (3) of this Act amended section 4(5) of the 1964 Act so as to require that the determination whether an accused is fit to plead and stand trial be made by the court and not a jury as section 4(4) of the 1964 Act had previously required.

7 Law Commission, *Unfitness to Plead Volume 1* (Law Com No 364, 2016).

8 ibid.

9 ibid [3.1-3.104]; Law Commission, *Unfitness to Plead Volume 2* (Law Com No 364, 2016) sections 3, 6, 32 and 34.

10 Law Commission, Equality, Capacity and Disability in Commonwealth Laws (ALRC Report 124, 2014) [7.4].

or institution.[11] A subsequent legislative amendment enacted in 2017, addressed the limited nature of the erstwhile orders available to South African courts and now bolster the rights to freedom and security of accused persons declared unfit to plead and stand trial. [12] And in India, following the ratification of the United Nations Convention on the Rights of Persons with Disabilities (UNCRPD) in 2007, and which has at aim "to promote, protect and ensure the full and equal enjoyment of all human rights and fundamental freedoms by all persons with disabilities, and to promote respect for their inherent dignity",[13] a new Indian Persons with Disability Act and Mental Health Act were respectively promulgated in 2016 and 2017 to better reflect the paradigm shift of mental disability from being a social welfare concern, to a human rights issue which underscores a "presumption of legal capacity, equality, and dignity".[14]

In stark contrast to this active and ongoing legislative agenda, is the comparable legislative framework for determining "fitness to plead and stand trial" in the Hong Kong Special Administrative Region. The laws of Hong Kong with regard to the determination of fitness to plead and stand trial, as well as the legal consequences following upon such a determination, have largely preserved the erstwhile position of the laws in England and Wales as it was established in the English Criminal Procedure (Insanity) Act 1964. Despite the provisions of the United Nations Convention on the Rights of Persons with Disabilities also finding application in Hong Kong, [15] the Hong Kong Law Reform Commission and the Legislative Council have yet to undertake a comprehensive legislative review and overhaul of this legislative framework. And arguably the most important and a necessary baseline for such legal reform, sufficient concern for and critical debate on the fair trial rights of accused persons unfit to plead and stand trial in the Hong

[11] De Vos NO v Minister of Justice and Constitutional Development 2015 (2) SACR 217 (CC).

[12] Criminal Procedure Amendment Act 4 of 2017.

[13] Convention on the Rights of Persons with Disabilities (A/RES/61/106, adopted on 13 December 2006, entered into force 3 May 2008) (CRPD) Preamble.

[14] Choudhary Laxmi Narayan and Deep Shikha 'Indian Legal System and Mental Health' (2013) 55:2 Indian Journal of Psychiatry 177; Richard M. Duffy and Brendan D. Kelly 'India's Mental Healthcare Act, 2017: Content, context, controversy' (2019) 62 International Journal of Law and Psychiatry 169.

[15] The United Nations Convention on the Rights of Person with Disabilities entered into force on 3 May 2008 and the People's Republic of China became a signatory to the Convention on 30 March 2007 and ratified the Convention on 1 August 2008. On the date of ratification, the People's Republic of China also indicated that the Convention shall apply to both the Special Administrative Regions of Hong Kong and Macau.

Kong Special Administrative Region must first amass in local (Hong Kong) legal discourse.

The primary aim of this article is to provide a comprehensive overview of Hong Kong laws and procedures relating to an accused person's competency to plead and stand trial. The legal test and related procedures for determining whether an accused person is possibly unfit to plead and stand trial will be considered, as well as the legal consequences that may ensue following such a determination. For every shortcoming identified in the article, possible solutions are recommended with reference to either the United Nations Convention on the Rights of Persons with Disabilities or to comparable legal development in the laws of England and Wales. Due to the historical and ongoing reliance of Hong Kong law reform and legal development on the laws of England and Wales, the discussion and analysis here are informed by the ongoing English legislative project for fitness to plead and stand trial, specifically the most recent 2016 Law Commission Report. This juxtaposition offers a particularly insightful analysis as the laws of Hong Kong with regard to fitness to plead and stand trial are still reflective of a rather archaic, legal paternalistic stance towards accused persons suffering from a mental, intellectual or cognitive disability, while the English legal development place increasing emphasis on a more supportive framework, which ensures the effective and equal participation of vulnerable accused in criminal proceedings.

II. The Legal Test for Determining Whether an Accused Person is 'Unfit to Plead and Stand Trial'

The first case in which criteria for determining fitness to plead were mentioned is *R v Dyson*,[16] a case involving a so-called "deaf-mute" who was determined to be "insane due to her inability to understand her right to challenge jurors".[17] This was followed in 1836, with the first formulation of a legal test to determine fitness to plead by Baron Alderson in *R v Pritchard*.[18] Of the inquiry into an accused assumed to be unfit to plead and/or stand trial, Baron Alderson stated that the following must be considered:

> [W]hether the prisoner is mute of malice or not; secondly, whether he can plead to the indictment or not; thirdly, whether he is of sufficient intellect to comprehend the course of proceedings on the trial, so as to make a proper defence—to know that he might challenge any of you to whom he may object—and to comprehend

[16] *R v Dyson* 173 ER 135; (1831) 7 Car & P 305.
[17] De Souza (n 3) 8.
[18] *R v Pritchard* (1836) 173 ER 135.

> the details of the evidence, which in a case of this nature must constitute a minute investigation. Upon this issue, therefore, if you think that there is no certain mode of communicating the details of the trial to the prisoner, so that he can clearly understand them, and be able properly to make his defence to the charge; you ought to find that he is not of sane mind. It is not enough, that he may have a general capacity of communicating on ordinary matters.[19]

It is important to remember that this test was formulated at a time when it was still impossible for an accused to give evidence in his own defence.[20]

Shortly after the decision in *Pritchard*, in the case of *R v Davies*,[21] the additional requirement that an accused must also be able to instruct his or her legal adviser was included. The test was subsequently interpreted to make it more consistent with the modern trial process. In England and Wales, for example, the present authoritative formulation of the test was articulated in *R v M (John)*[22] where the judge directed the jury that they should find the accused unfit to plead if any one or more of the following was beyond the capacity of the accused: understanding the charge(s), deciding whether to plead guilty or not, exercising the right to challenge jurors, instructing solicitors and/or barristers, following the course of the proceedings and giving evidence in his or her own defence.[23] And more recently in *R v Orr*,[24] Judge Macur for the English Court of Appeal held that "fitness to plead" is more aptly identified as "'fitness to participate in the trial process', since 'the supposed disability' can be determined at any stage" of the trial proceedings.[25] For example, the appellant, in this case, suffered from depression and while he was fit to enter a plea and to participate in some of the criminal proceedings relating to money laundering charges against him, his mental health deteriorated, and he was unable to finish his evidence-in-chief and unable to undergoing cross-examination.[26] Yet, the trial judge allowed for the proceedings to conclude but directed that the appellant should not undergo cross-examination and gave strict instructions to the prosecution not to refer in closing argument to any subject which the appellant had

[19] ibid, 135; *R v Sharp* [1960] 1 QB 357.
[20] *R v Orr* [2016] 4 WLR 132 6.
[21] *R v Davies* (1853) 3 Car & Kir 328, 175 ER 575.
[22] *R v M (John)* [2003] EWCA Crim 3452 [2003] All ER (D) 199.
[23] Note that in 2004, with the enactment of the Domestic Violence, Crime and Victims Act 2004 in England and Wales, the law changed to a judge-only determination of an accused's fitness to plead and stand trial. De Souza (n 3) 8. Also see *R v Walls* [2011] EWCA Crim 443; [2011] 2 Cr App R 6.
[24] *R v Orr* [2016] 4 WLR 132.
[25] ibid 135-6.
[26] ibid 134-5.

not been cross-examined on. At the heart of the appeal was whether the appellant became unfit to be tried during the trial or whether he merely became unable to give evidence and be cross-examined.[27] Judge Macur held that once the issue of "fitness to plead" had been raised, it must be determined. In this case, the trial judge, in finding that the appellant had been fit to participate in his trial up to the point of cross-examination, implicitly also found that the appellant subsequently became unfit to fully participate fully in his trial according to the *Pritchard* criteria.[28]

This legal test, as formulated in *R v Pritchard*,[29] was also adopted in Hong Kong law in *R v Leung Tak-Choi*,[30] where it was held that disability in this context "refers to the lack of ability to understand the charge against him, to give instructions to his lawyers, to challenge the jurors, to understand the evidence against him, and to give evidence in defence."[31] The mere fact, therefore, that an accused suffers from a mental illness or an intellectual or cognitive impairment or some other disability will not be sufficient for a finding that the accused is unfit to plead and stand trial. It must be shown that this disability or impairment affects the ability of the accused to meaningfully and effectively participate in the criminal proceeding against him or her.[32] Of this test, the English Law Commission noted the following:[33]

- Uncertainty exists about the formulation of the text, its scope and proper application and is, as a result, not widely and consistently applied.
- The test focuses too heavily on the intellectual ability of an accused and fails to consider other aspects of mental illness and other

[27] ibid 136.

[28] ibid 136-7. Note that in terms of section 35(1)(b) of the Criminal Justice and Public Order Act 1994 a court under the laws of England and Wales may not draw an adverse inference from an accused person's silence where it appears to the court that the physical or mental condition of the accused makes it undesirable for him to give evidence.

[29] *R v Pritchard* (1836) 173 ER 135.

[30] *R v Leung Tak-Choi* [1995] HKCFI 202; HCCC 457/1994 (26 June 1995).

[31] ibid [31].

[32] Julio Arboleda-Florez 'Fitness to Stand Trial – Is it Necessary' (1982) 26:1 International Journal of Offender Therapy and Comparative Criminology 43, 43-44, quoting from Norwood W. East, *An introduction to forensic psychiatry in criminal courts* (London: J.A. Churchill 1927).

[33] Law Commission, *Unfitness to Plead Summary* (Law Com No 364, 2016) paras 1.42-1.47; Law Commission, *Unfitness to Plead Volume 1* (Law Com No 364, 2016) [3.11]. Also see Amar Shah 'Making Fitness to Plead Fit for Purpose' (2012) 1 International Journal of Criminology and Sociology 176.

conditions which may interfere with an accused's ability to engage effectively in the trial process.

- No explicit consideration is given to an accused's ability to make decisions required of him or her during the criminal proceeding.

- There is a general lack of clarity over the correlation between an accused being unfit to plead, and the fair trial guarantee that an accused must participate effectively in the proceedings against him or her.

- Finally, the Law Commission also noted that "[t]he current test and procedures do not allow a defendant who would otherwise be unfit for trial, but who clinicians consider has the capacity to plead guilty, to do so. This may unnecessarily deny the defendant his or her legal agency. It is also liable to undermine victim confidence in the system and deny the court the opportunity to impose sentence where appropriate."[34]

These comments by the English Law Commission are of equal importance to the Hong Kong statutory scheme given its shared origin in the 1836 decision in *R v Pritchard*.[35]

The English Law Commission emphasised two important aspects with regard to a modern legal test for determining unfitness to plea and stand trial: First, it was held that the current terminology in terms of "a finding of 'disability' such that the defendant is deemed 'unfit to plead', is outdated…[a]nd risks labelling a defendant in a way that may be objectionable to him or her and to others affected by the proceedings."[36] Second, that a reformulated test should rather focus on assessing an accused person's ability to participate effectively in his or her trial.[37] Requiring "effective participation" of an accused person in a criminal trial was described by the European Court of Human Rights as "implicit in the very notion of an adversarial procedure",[38] and inclusive of the following: the accused must have "a broad understanding of the nature of the trial process and of what is at stake for him or her, including the significance of any penalty which may be imposed. It means that he or she, if necessary, with the assistance of, for example, an interpreter, lawyer, social worker or friend, should be able to understand the

[34] Law Commission, *Unfitness to Plead Summary* (Law Com No 364, 2016) [1.15].
[35] *R v Pritchard* (1836) 7 C & P 303, 173 ER 135.
[36] Law Commission, *Unfitness to Plead Volume 1* (Law Com No 364, 2016) [3.34].
[37] ibid.
[38] *Stanford v United Kingdom* App No 16757/90 (23 February 1994) [26].

general thrust of what is said in court. The defendant should be able to follow what is said by the prosecution witnesses and, if represented, to explain to his own lawyers his version of events, point out any statements with which he disagrees and make them aware of any facts which should be put forward in his defence."[39] An accused must therefore be able to maintain an active level of involvement throughout the trial, and must also be able to make relevant decisions during the course of the trial itself.[40]

Thus, in steering away from the concepts of "disability" and "unfit to plead", the English Law Commission proposed a legal test of capacity, based on a list of relevant abilities which taken together must be sufficient to enable effective participation of the accused in the legal proceedings against him or her. These relevant abilities include, *inter alia*, the ability to understand the nature of the charge, the nature and purpose of the evidence adduced, as well as the nature of the trial process and the legal consequences following a conviction, and the ability to give instructions to a legal representative.[41] This new legal test also incorporates explicitly the need for an accused person to be able to make key decisions, like whether to plead guilty or not guilty and whether to testify at trial. It is prescribed that this ability to make key critical decisions is dependent on the accused person's ability to understand information relevant to the making of the decision, the ability to retain that information, and to use and weigh this information in making an informed decision, and to, ultimately, communicate this decision in court.[42]

In addition, the English Law Commission also emphasised the importance of protecting an accused person's fundamental right to legal autonomy and recognised the significant impact that a finding of unfit to plead and stand trial has:

> "[s]uch a curtailment of an individual's access to justice should only occur where absolutely necessary in the circumstances, and where it is essential to protect other fundamental rights of the defendant, in this situation the right to a fair trial…[i]t is in the interests not only of the accused but of all those affected by an alleged offence that all those who can fairly be tried for an offence in the usual way should be".[43]

The Commission consequently recommended that a separate legal test of capacity to plead guilty be included in statute and that this test only be applied in

[39] *SC v United Kingdom* (2005) 40 EHRR 10 (App No 60958/00) [29].

[40] Law Commission, *Unfitness to Plead Volume 1* (Law Com No 364, 2016) [3.21].

[41] Law Commission, *Unfitness to Plead Volume 2* (Law Com No 364, 2016) section 3(4)(a)-(d) of the Criminal Procedure (Lack of Capacity) Bill.

[42] Law Commission, *Unfitness to Plead Volume 2* (Law Com No 364, 2016) section 3(4)(e)-(g) and 3(5)(a)-(d) of the Criminal Procedure (Lack of Capacity) Bill. Also see Law Commission, *Unfitness to Plead Volume 1* (Law Com No 364, 2016) para 3.2.

[43] Law Commission, *Unfitness to Plead Volume 1* (Law Com No 364, 2016) para 3.40.

cases where an accused has been found unfit to plead and stand trial and where there is expert opinion considering whether the accused nonetheless has the capacity to plead guilty. It is also required that an accused specifically applies to the court to invoke this provision and determine the issue.[44]

These two proposed new legal tests[45] certainly provide for greater legal certainty, and ultimately also greater consistency in the determination of whether an accused person is fit to plead and to stand trial. The two tests are, furthermore, consonant with the provisions of the UN Convention of the Rights of Persons with Disabilities (which also applies in Hong Kong) in ensuring that all accused persons are treated equally before the law and that those persons who suffer from a disability receive the necessary assistance and opportunity to also exercise their legal autonomy also.[46] Finally, it can be noted that the two tests proposed by the English Law Commission for determining capacity to plead and stand trial give full consideration to the notion of "effective participation" throughout the various stages of a criminal proceeding and ultimately also provide for the possibility to enter a guilty plea.

The rationale and design of the ongoing English law reform with regard to the legal test for determining fitness to plead and stand trial, and as set out above, certainly lay bare the shortcomings of the old common law test originally formulated in *R v Pritchard*,[47] and adopted in Hong Kong law in *R v Leung Tak-Choi*.[48] This common law test not only falls short in adequately assessing the ability of an accused person to participate effectively and meaningfully in his or her trial, but it is also not conducive of promoting equal protection before the law. Legal reform and development in this regard is therefore certainly necessary; a Hong Kong statutory test for determining fitness to plead and stand trial will not only guarantee the fair trial rights of accused persons suffering from a mental, intellectual or a cognitive disability, it will also allow for greater legal certainty and consistency in the determination of fitness to plead and stand trial.

To this end, shortcomings in ancillary legislation in Hong Kong also need to be addressed. For example, while the Mental Capacity Act 2005 of England and Wales provides for a legal definition of the concept "capacity", no comparable

[44] Law Commission, *Unfitness to Plead Volume 1* (Law Com No 364, 2016) para 3.2. Law Commission, *Unfitness to Plead Volume 2* (Law Com No 364, 2016) section 6 of the Criminal Procedure (Lack of Capacity) Bill.

[45] Law Commission, *Unfitness to Plead Volume 2* (Law Com No 364, 2016); Criminal Procedure (Lack of Capacity) Bill s 33, 35.

[46] United Nations Convention on the Rights of Persons with Disabilities (CRPD) (n 13) articles 12-13.

[47] *R v Pritchard* (1836) 7 C & P 303, 173 ER 135.

[48] *R v Leung Tak-Choi* [1995] HKCFI 202; HCCC 457/1994 (26 June 1995).

definition exists in Hong Kong law. In Hong Kong law, "mental incapacity" is defined in the Mental Health Ordinance as "mental disorder or mental handicap" and these two concepts are respectively described as follows: "Mental disorder means (a) mental illness; (b) a state of arrested or incomplete development of mind which amounts to a significant impairment of intelligence and social functioning which is associated with abnormally aggressive or seriously irresponsible conduct on the part of the person concerned; (c) psychopathic disorder; or (d) any other disorder or disability of mind which does not amount to mental handicap." And "mental handicap" refers to a "sub-average general intellectual functioning with deficiencies in adaptive behaviour, and 'mentally handicapped' shall be construed accordingly." These definitions for "mental incapacity" and "mental handicap" under Hong Kong law are clearly inadequate in terms of our contemporary understanding of the full extent of a person's mental, intellectual and cognitive abilities. The existing Hong Kong definitions in this regard rather pathologise and medicalise human/personal capacity by confining the definitions to illness, disorders, impairments, and handicaps. Such excessive pathologising of a basic human capability, i.e. the capacity to make decisions, not only contributes to the stigma surrounding mental illness, but also results in a distorted view of the vast continuum on which a person's capacity and incapacity, or ability and inability to make decisions, can exist, and the fact that any determination in this regard may fluctuate with time, or with regard to particular decisions at material times.

For example, in the preamble of the United Nations Convention on the Rights of Persons with Disabilities it is recognised explicitly that "[d]isability is an evolving concept and that disability results from the interaction between persons with impairments and attitudinal and environmental barriers that hinders their full and effective participation in society on an equal basis with others".[49] It is, furthermore, recognised that discrimination against any person on the basis of disability is a violation of the inherent dignity and worth of the human person and that a wide diversity of persons with disabilities exist.[50] A non-medicalised articulation of capacity in the legal sense can be found in the English Mental Capacity Act 2005 Cap 9, where a person who lacks capacity is described in terms of that person lacking "capacity in relation to a matter if at the material time he is unable to make a decision for himself in relation to the matter because of an impairment of, or a disturbance in the functioning of, the mind or brain."[51] For the purpose of this definition it does not matter whether the impairment or disturbance is permanent or temporary, nor can this lack of capacity be established

[49] ibid Preamble (e).
[50] ibid Preamble (h)-(i).
[51] Mental Capacity Act, s 2(1).

merely by reference to age, appearance, a condition, or any aspect of behaviour.[52] Likewise, the inability to make decisions is defined specifically in the English Mental Capacity Act 2005 Cap 9 in terms of a person being unable "(a) to understand the information relevant to the decision, (b) to retain that information, (c) to use or weigh that information as part of the process of making the decision, or (d) to communicate his decision (whether by talking, using sign language or any other means)."[53]

While there has been no indication of legal reform in Hong Kong for a new legal test for determining fitness to plead and stand trial, the shortcomings of the Hong Kong legal definition of "capacity" were remarked upon by the Hong Kong Law Reform Commission in August 2006 with its report titled *Substitute Decision-making and Advance Directives in Relation to Medical Treatment*.[54] No meaningful legal reform has taken place to date. The updating and alignment of a legal test and ancillary legislation dealing with accused persons suffering from a cognitive, mental, or intellectual disability in the criminal justice process are, however, imperative and should be considered in conjunction with the procedural aspects relating to this far-reaching pronouncement on an accused person's abilities at trial.

III. The Procedure for Accurate and Efficient Identification of Accused Persons Unable to Meaningfully and Effectively Participate in Their Trial

Section 75 and 75A of the Criminal Procedure Ordinance of Hong Kong provide for the procedure to determine whether an accused person at trial suffers from an incapacity or disability that affects their fitness to plead and be tried. These provisions are derived from the English Criminal Procedure (Insanity) Act 1964 and save for minor amendments, still reflect the law of England and Wales in this regard before the legislative revisions brought about in the Criminal Procedure

[52] Mental Capacity Act, s 2(2)-(3).

[53] Mental Capacity Act, s 3(1).

[54] Law Commission, Substitute Decision-making and Advance Directives in Relation to Medical Treatment (HKLRC, August 2006).

(Insanity and Unfitness to Plead) Act 1991,[55] and the Domestic Violence, Crime and Victims Act 2004.[56]

According to section 75(1) of the Criminal Procedure Ordinance, the question of whether an accused is "under a disability…[that] would constitute a bar to his being tried", arises "at the instigation of the defence or otherwise". There is no duty on a presiding officer to raise the issue of an accused's competence *mero motu* or where there is merely a possibility of some disability or lack of ability, and it is generally accepted that a presiding officer will only intervene if a substantial question as to the defendant's trial competency is raised; "a heavy burden and responsibility [therefore] rests on legal representatives—especially those for the defence—to make known the suspicious behaviours of defendants and that might suggest incompetency."[57] The defence must ultimately provide clear and convincing evidence that the accused is unfit to plead and stand trial, and section 75A details the procedure for such a finding to be made.[58]

Yet, effectively and accurately identifying instances where the accused may be unfit to plead and stand trial is not always an easy feat. The appellant in *HKSAR v Choi Yiu Wai David*[59] appealed against his conviction on a charge of theft contrary to section 9 of the Theft Ordinance on the basis that his guilty plea was a nullity. [60] The appellant explained that his legal representation under the Duty Lawyer Scheme at trial was ineffective as he suffered from a delusional and obsessive-compulsive disorder and was not taking his psychotropic drugs at the time of trial. This, he submitted, resulted in him being overly anxious and nervous during the conference with his duty lawyer. [61] It was also during this pre-trial conference that the appellant experienced some of the recurrent thoughts that often consumed his mind; he felt as if his head was shrinking and that he had to check the circumferences of his head to confirm that his head was not shrinking.[62] On appeal before the Hong Kong Court of First Instance this case was described

[55] The amendments in this Act was prompted by the Report of the Committee on Mentally Abnormal Offenders (Cmnd 6244, 1975), generally referred to as The Butler Report, and included, inter alia, for a mandatory hearing of the facts of the case once an accused has been found to be unfit to plead (section 4A).

[56] Section 22(1) to (3) of this Act amended section 4(5) of the 1964 Act so as to require that the determination whether an accused is fit to plead and stand trial be made by the court and not a jury as section 4(4) of the 1964 Act had previously required.

[57] *R v Keung Sai-chung* (1986) HKLR 838 (CA); Samuel Adjorlolo, Heng Choon and Oliver Chan "Determination of Competency to Stand Trial (Fitness to Plead): An Exploratory Study in Hong Kong" (2017) 24:2 Psychiatry, Psychology and Law 205, 208.

[58] *R v Podola* [1960] 1 QB 235.

[59] *HKSAR v Choi Yiu Wai David* [2011] HKCFI 1847; HCMA 459/2011 (16 December 2011).

[60] ibid [1]-[2].

[61] ibid [17].

[62] ibid [12].

as one where the appellant's mental illness was such that a layman would not notice it unless the appellant mentioned it. Judge Barnes found that the appellant did not have a fair trial as his outward appearance and behaviour would not have alerted the magistrate or the duty lawyer of his mental condition, "it is a fact that he was not a man free of any mental illness at the time. As such, there is a real likelihood that the appellant's mental process was affected to such an extent that his mind did not truly go with his act, making his act of pleading guilty not a true act."[63]

While the trial court, prosecution, and legal counsel in *HKSAR v Choi Yiu Wai David* had failed to accurately identify the appellant as possibly unfit to plead, the opposite happened in *HKSAR v Chan Shu Hung*,[64] where the magistrate, in assuming that the accused was not fit to stand trial, invoked section 51 of the Mental Health Ordinance and remanded the accused to a mental hospital for observation. The accused had no legal representation at trial, he elected not to give evidence in his defence, and he also did not call any witness(es) in support of his case. In denying that he was responsible for criminal damage contrary to section 60(1) of the Crimes Ordinance, the accused presented the magistrate with a letter in which he wrote out his defence. On numerous occasions during the trial the accused requested that the magistrate read this letter but the magistrate ignored these requests and took the view, shortly before the prosecution closed its case, that the accused was "speaking in a somewhat incoherent and confused manner".[65] It was based on this observation that the magistrate decided to remand the accused for observation in a mental hospital.[66] Having read the transcript and having listened to the recording of the trial, Judge Barnes for the Hong Kong Court of First Instance was unable to detect any instance of the appellant speaking "'in an incoherent and confused manner' or hear anything which sounds like the appellant having any mental problem."[67] Judge Barnes also agreed with the defence that the trial magistrate did not have a good understanding of the defence case and that he may, for this reason, have held the opinion that the appellant spoke in an incoherent and confused manner.[68]

In terms of the applicable procedure, it can be noted that whether an accused person is fit to plead and stand trial usually falls to be determined upon arraignment and this can, of course, occur at the level of the Magistrate's Court, the District Court, or the Court of First Instance. While the procedure for making

[63] ibid [27].
[64] *HKSAR v Chan Shu Hung* [2011] HKCFI 1853; [2012] 2 HKLRD 424; HCMA 425/2011 (20 December 2011).
[65] ibid [11].
[66] ibid [11].
[67] ibid [32].
[68] ibid [32].

the determination in the Court of First Instance are detailed in the Criminal Procedure Ordinance, very little statutory guidance exists at the Magistrate's Court and District Court level. The discussion here focuses on the procedural aspects at the level of the Court of First Instance. With regard to the procedure at Magistrate's Court and District Court level it can be noted that many of the provisions in the Mental Health Ordinance refer to both a "court or magistrate"[69] or to "the District Judge or magistrate",[70] and it can therefore be inferred that many of the orders in terms of the Mental Health Ordinance can also be made by a Magistrate or District Court Judge.[71]

Jury determination

At the Court of First Instance, a jury is tasked to decide, as a question of fact, whether an accused person is fit to plead and stand trial. A jury so tasked will make its decision based on the written or oral evidence of two or more registered medical practitioners of whom no less than two shall be psychiatrists on the Specialist Register established under section 6(3) of the Medical Registration Ordinance Cap 161.[72] If the jury decides that the accused person is indeed mentally fit and able, the trial will proceed but before another jury. Likewise, where the question as to the fitness of an accused person to stand trial falls to be determined at any later time after arraignment it is usually also decided by a jury other than the jury tasked with deciding on the guilt or innocence of an accused.[73] In all other instances the question as to the fitness of the accused to stand trial will be determined by the same jury tasked with deciding on the guilt or innocence of the accused.[74]

Where it is determined by a jury that an accused person is not fit to stand trial, the trial shall not proceed or further proceed and the jury shall determine whether the accused did the act or omission with which he or she is charged, based on such evidence as may be adduced or further adduced by the prosecution or by a person appointed by the court to put the case for the defence. A factual finding can subsequently follow as to whether the accused person did or did not do the act or omission so charged. Where the finding is negative, the jury shall return a verdict

[69] See, for example, sections 44A, 44D, 44E, 44F, 44I, 45, 46, 47 etc.

[70] See, for example, sections 31, 32, 36, 62, 68, 71 etc.

[71] In section 2 of the Mental Health Ordinance, 'court' is defined as the Court of First Instance and any judge of the Court of First Instance.

[72] The Criminal Procedure Ordinance Cap 221, Section 75(5).

[73] The Criminal Procedure Ordinance Cap 221, Section 75(4)(a)(i) and 75(4)(b)(i). Note that it is possible for the court to also direct otherwise, i.e. that the same jury decides on both the fitness of the accused to plead and stand trial as well as the guilt or innocence of the accused.

[74] The Criminal Procedure Ordinance Cap 221, Section 75(4)(a)(ii) and 75(4)(b)(ii).

of acquittal as if the trial had proceeded to a conclusion on that particular count.[75] A positive finding, on the other hand, remains a mere factual determination and does not constitute a guilty verdict in the conventional sense. It therefore, remains possible for such an accused person, at a later stage once his or her mental capabilities have improved, to contest the factual finding in the same manner as it is possible for an accused to be tried at a later stage in respect of the act or omission so charged.[76] Previously, the question of an accused's guilt or innocence in the event that the accused was found to be unfit to plead and stand trial, was postponed until such time as the accused had regained the requisite faculties to meaningfully participate in his or her own defence.[77] It was explained in *R v Leung Tak-Choi* that the mandatory hospital order in terms of the erstwhile version of section 76 of the Criminal Procedure Ordinance Cap 221 was warranted for this very reason, as the accused had to be detained and had to receive compulsory medical treatment and care for his or her guilt or innocence to be decided at a later stage, when he or she was fit to be tried again.[78] The advantage of the factual determination as now contemplated in section 75A of the Ordinance is that it provides for closure in the criminal proceeding and even a full verdict of acquittal where the jury finds that the accused person did not commit the act or omission as alleged. However, it must also be noted that while a positive finding for whether the accused had committed the act(s) or omission(s) so charged, will only amount to a factual finding and not a *true* guilty verdict, it nonetheless is a pronouncement (with consequences) on the acts or omissions of an accused person who is at that moment unable to participate meaningfully and effectively in his own defence. For example, in *R v Orr*[79] it was argued on behalf of the appellant that a positive factual finding, although not a conviction, remains something that is particularly hard to bear "because it is a form of stigma, whatever the law says about it".[80] To make a positive factual finding may furthermore be difficult where failure to give a satisfactory explanation is an essential ingredient to the offence.[81]

The correct interpretation of these provisions on the empanelling of a jury for the purpose of deciding whether an accused is fit to plead and stand trial was considered in *HKSAR v Ng Mei Lan* The applicant, in this case, was charged with

[75] The Criminal Procedure Ordinance Cap 221, Section 75A(1).

[76] The Criminal Procedure Ordinance Cap 221, Section 76(4).

[77] See generally *R v Leung Tak-Choi* [1995] HKCFI 202; HCCC 457/1994 (26 June 1995).

[78] *R v Leung Tak-Choi* [1995] HKCFI 202; HCCC 457/1994 (26 June 1995) [31]. The amended section 76 of the Criminal Procedure Ordinance Cap 221 now provides for a number of orders available to a court once it has been determined that an accused person is unfit to plead and stand trial. Section 76 and the various orders available to courts will be considered in the second article.

[79] *R v Orr* [2016] 4 WLR 132.

[80] ibid [3].

[81] For example, failure to give a reasonable explanation on suspicion of possession of stolen goods.

manslaughter contrary to common law and punishable under section 7 of the Offences Against the Person Ordinance Cap 212, and arson with intent contrary to sections 60(2) and (3) and 63(1) of the Crimes Ordinance Cap 200.[82] A jury was empanelled pursuant to section 75 of the Criminal Procedure Ordinance Cap 221 to determine whether the applicant was fit to be tried.[83] This jury unanimously determined that the applicant was not fit to be tried and the same jury then proceeded to find that the acts alleged had been proved pursuant to section 75A of the Ordinance.[84] The applicant was subsequently ordered to be detained in Siu Lam Psychiatric Centre under section 76(2)(a)(i) and Schedule 4 of the Criminal Procedure Ordinance Cap 221.[85] At issue on appeal was whether section 75A of the Criminal Procedure Ordinance Cap 221 required another jury to be empanelled to determine whether the applicant did the acts charged or whether the same jury having decided that the applicant was unfit to plead and stand trial could also make the factual determination as to whether the applicant did or did not do the act or omission so charged.[86] Judges Stuart-Moore, Wright, and Saw for the Hong Kong Court of Appeal held that where a jury empanelled in the Court of First Instance is tasked with deciding whether an accused is fit to stand trial, another jury must be empanelled for the trial.[87] The obvious reason for this requirement is that the first jury may have heard evidence from experts and possibly also the accused him- or herself, and also evidence about the nature and the circumstances of the alleged offence that may cloud their judgment at trial.[88] While the issue of an accused's fitness to stand trial arise after arraignment, the situation is different; "[u]p to that point the trial will have proceeded in the normal way. Evidence more prejudicial than probative will have been excluded. Evidence irrelevant to the matters in issue will not have been placed before the jury".[89] Thus, where the issue of an accused's fitness to stand trial arise only after arraignment, the wording of section 75A of the Criminal Procedure Ordinance Cap 221 is clear that it is the same jury that makes a positive determination on the question of fitness to stand trial, that will then also consider the subsequent issue whether that accused had committed the act or

[82] *HKSAR v Ng Mei Lan* [2009] HKCA 44; [2009] 3 HKLRD 193; [2009] 3 HKC 277; CACC 149/2008 (12 February 2009) [1].
[83] ibid [2].
[84] ibid [3], [6].
[85] ibid [8].
[86] ibid [4].
[87] The Criminal Procedure Ordinance Cap 221, Section 75(4)(a); *HKSAR v Ng Mei Lan* (n 38) para [18].
[88] HKSAR v Ng Mei Lan (n 38) [19].
[89] ibid [19].

omission in question.[90] Another (new) jury will only be empanelled if the defendant is determined not to be under any disability.[91]

A Hong Kong court faced with the question of an accused's fitness to be tried may also postpone the matter until a time up to the opening of the case for the defence. However, where the question does not arise until after the jury has returned a verdict of acquittal on the count(s) for which the accused person is tried, the question of that accused person's fitness to stand trial shall not be determined.[92] The question can furthermore arise on appeal against a conviction, as per section 83J of the Criminal Procedure Ordinance Cap 221. Section 83L of the Ordinance will then apply and the provisions of this section are, in essence, similar to the provisions relating to the question of fitness to stand trial in a court a quo and which will be set out below.

The most significant difference in the procedure whereby it is to be determined whether an accused person is fit to plead and stand trial under Hong Kong law and the laws of England and Wales relates to the role of the jury. In 2005, by way of section 22(2)-(3) of the Domestic Violence, Crime and Victims Act 2004 (c. 28), section 4(5)-(6) of the English Criminal Procedure (Insanity) Act 1964 was amended to exclude the jury from the determination whether an accused person is fit to plead and stand trial. Lord Auld in his 2001 Review of the Criminal Courts of England and Wales explained that:

> In the majority of cases the jury's role on the issue of unfitness to plead is little more than a formality because there is usually no dispute between the prosecution and the defence that the defendant is unfit to plead. However, the procedure is still cumbrous, especially when the issue is raised, as it mostly is, on the arraignment, because it can then require the empanelling of two juries. More importantly it is difficult to see what a jury can bring to the determination of the issue that a judge cannot. He decides similar questions determinative of whether there should be a trial, for example, whether a defendant is physically or mentally fit to stand or continue trial in applications to stay the prosecution or for

[90] ibid [22].

[91] ibid [24], [26]-[28]; It is interesting to note that the Hong Kong provisions in this regard depart from the original English legislation on which the Hong Kong provisions were modelled: In terms of sections 2 and 4A of the Criminal Procedure (Insanity and Unfitness to Plead) Act 1991, where a jury found an accused to be unfit to stand trial, a separate jury had to be empanelled to determine whether that accused did the acts or omissions so alleged.

[92] The Criminal Procedure Ordinance Cap 221, Section 75(2).

discharge of the defendant.[93]

Moreover, even where the determination is made by a presiding judicial officer alone, the English Law Commission described the process as time-consuming, leading to substantial delays, and causing uncertainty and anxiety to complainants, witnesses, and the defendant.[94] This is all the more so where a jury is tasked with determining whether an accused is fit to plead and stand trial. In Hong Kong, however, and as was indicated above, differing procedures currently exist with regard to the determination in Magistrate's Courts and the District Court, versus the procedure in the Court of First Instance. While the determination stands to be made by the presiding judicial officer in a Magistrate's Court and the District Court, a jury is tasked with making the determination in the Court of First Instance.

It is submitted that the differing procedures currently in place under Hong Kong law for determining whether an accused person is fit to plead and stand trial are unnecessary, breed inconsistency, and are particularly time-consuming in the Court of First Instance where a jury is tasked to make this determination. Most important, is that the accumulation of knowledge and awareness of mental health and cognition, as well as the interminable efforts to develop procedures that guarantee and safeguard the fair trial rights of accused persons, have rendered the determination whether an accused person is fit to plead and stand trial a legal-technical matter that depends on the correct application of the law to a particular factual set, and it can no longer be left for a jury to decide only as a matter of fact. It is for this reason also that it is a requirement in both the laws of England and Wales as well as in Hong Kong that the consequences that follow from such a determination be informed by the written or oral evidence of two or more registered medical practitioners that meet the requirements of the respective statutes.[95] Moreover, in England and Wales, the Law Commission recommended that all members of the judiciary and all legal practitioners engaged in criminal proceedings receive training in understanding and identifying participation and communication difficulties on the part of accused persons. This, it was said, "would improve accurate and timely identification of participation difficulties, reducing delays to proceedings and the uncertainty and anxiety caused to complainants and witnesses where the defendant's participation difficulties are raised at the last

[93] Law Commission, *Unfitness to Plead* (Law Com CP 197, 2010) [2.33].

[94] Law Commission, *Unfitness to Plead Summary* (Law Com No 364, 2016) para 1.47; Law Commission, *Unfitness to Plead Volume 1* (Law Com No 364, 2016) Chapter Four.

[95] See the English Criminal Procedure (Insanity) Act 1964, section 4(6); and the Hong Kong Criminal Procedure Ordinance Cap 221, section 75(5).

minute".[96] This acknowledgement that appropriate training for legal professionals is imperative for the just administration and determination of whether an accused person is fit to plead and stand trial further supports the submission that this determination can no longer be for a lay jury to decide based on evidence, argument, and inferences alone.

IV. The Consequences Following a Determination that an Accused Person is Unfit to Plead and Stand Trial

Section 76 of the Criminal Procedure Ordinance Cap 221 applies whenever a special verdict was returned under section 74 of the Criminal Procedure Ordinance Cap 221 finding the accused not guilty by reason of insanity, or where it was found under sections 75 and 75A of the Ordinance that the accused is unfit to stand trial, but that the accused person did in fact commit the offence(s) with which he or she is being charged.[97] Then, in terms of section 76(2)(a), a court may admit the accused person to a mental hospital or a Correctional Psychiatric Centre if it is satisfied on the written or oral evidence of two or more registered medical practitioners (of whom not less than two are registered under the Medical Registration Ordinance Cap 161 as psychiatrists) that it is necessary in the interests or the welfare of the accused person, or for the protection of other persons to have the accused so admitted. In addition to an order admitting the accused to a mental hospital or a Correctional Psychiatric Centre,[98] provision is also made for a guardianship order under Part IIIA of the Mental Health Ordinance Cap 136,[99] a supervision and treatment order under Part IIIB of that Ordinance, or an order for the absolute discharge of the accused.[100] These alternative orders are not available to a court where the special verdict returned in terms of section 74 of the Ordinance or the finding in terms of sections 75 and 75A of the Ordinance relate to an offence for which the sentence is fixed by law.[101] For such cases, a hospital order must be imposed. Where a jury has decided that an accused is unfit to stand trial and that the accused person has also not committed the act or omission so charged, a verdict of not guilty will be returned and the presiding judicial officer may also make an appropriate order which is in the best interests and welfare of the acquitted person as well as for the protection of society in terms of the

[96] Law Commission, *Unfitness to Plead Summary* (Law Com No 364, 2016) [1.35]; Law Commission, *Unfitness to Plead Volume 1* (Law Com No 364, 2016) [2.21-2.30].

[97] The Criminal Procedure Ordinance Cap 221, Section 76(1)(a)-(b).

[98] ibid Section 76(2)(a).

[99] *HKSAR v Cheung Kam Yau* [2017] HKCFI 507; HCCC 413/2016 (22 March 2017).

[100] The Criminal Procedure Ordinance Cap 221, Section 76(2)(b)(i)-(iii).

[101] ibid Section 76(3).

provisions of the Mental Health Ordinance Cap 136. In the least serious of cases a court may order for the absolute discharge of such an accused.[102]

Any order made in terms of section 76 of the Criminal Procedure Ordinance Cap 221 or the provisions of the Mental Health Ordinance Cap 136 is appealable either in terms of sections 83M and 83N of the Criminal Procedure Ordinance Cap 221 or before the Mental Health Tribunal[103] or the Guardianship Board,[104] as the case may be. Interesting aspects with regard to the appeal procedure include the following: In terms of section 83N of the Criminal Procedure Ordinance Cap 221, it is provided that where the question of fitness to stand trial was determined later than on arraignment, and the Court of Appeal is of the opinion that the case was one in which the accused person should have been acquitted before the question of fitness to be tried was considered, the Court of Appeal shall not only quash the conviction and finding of unfitness to stand trial, but the Court of Appeal shall also direct that a verdict of acquittal be recorded but not a verdict of not guilty by reason of insanity.[105] It is also possible for the Chief Executive to refer a case where the accused was found not guilty by reason of insanity or unfit to stand trial because of a disability to the Court of Appeal. Such referrals are then also regarded and dealt with as if it was an appeal against the findings.[106]

It is evident from the above, that courts, upon finding that an accused person is unfit to plead and stand trial, have a relatively wide range of options available in terms of section 76 of the Criminal Procedure Ordinance Cap 221. Moreover, any determination and order so made are appealable and subject to the scrutiny and interjection of other executive organs like the Mental Health Tribunal, the Guardianship Board, or the Chief Executive. These legislative protections have not always been in place in Hong Kong law to the extent that they are today.[107] In terms of the previous version of section 76 of the Criminal Procedure Ordinance Cap 221, once a determination had been made that an accused person is unfit to plead and stand trial, the court was obliged to impose a mandatory hospital order, without this order being conditioned upon medical evidence, and to postpone the proceedings until such time as the accused had regained the requisite faculties

[102] ibid Section 72(2)(b)(iii); *HKSAR v Cheung Kam Yau* [2017] HKCFI 507; HCCC 413/2016 (22 March 2017).

[103] The Mental Health Ordinance Cap 136, Section 59A.

[104] ibid Section 59J and 59K.

[105] The Criminal Procedure Ordinance Cap 221, Section 83N(2).

[106] ibid Section 83P.

[107] See *R v Leung Tak-Choi* [1995] HKCFI 202; HCCC 457/1994 (26 June 1995). This case was decided under the previous version of section 76 of the Criminal Procedure Ordinance Cap 211 when the imposition of a hospital order was still mandatory once it has been found that an accused is unfit to plead and stand trial.

to meaningfully participate in his or her own defence.[108] Under this regime, an accused person found to be unfit to plead and stand trial was, in a sense, worse off than "a person setting up the defence of insanity because that kind of person would at least have gone through a trial whereas a person under disability would not have such an opportunity."[109] This position changed with the enactment of section 75A which now provides for a factual determination to be made as to whether an accused did or did not commit the act(s) or omission(s) so charged once it has been found that the accused is unfit to plead and stand trial. This factual determination has rendered the previous mandatory hospital order nugatory and justifies the wider range of options now available to courts.

Yet, although some safeguards therefore exist in the statutory scheme for fitness to plead and stand trial in Hong Kong, and a wider range of options are now available to courts upon a finding that an accused is unfit to plead and stand trial, the question can also be asked here, with regard to the legal consequences following a determination of unfit to plead and stand trial, whether such accused persons are afforded *equal* participation before and *protection* of the law. Equal participation before the law requires disability neutral rules and procedures, as well as the necessary support for vulnerable accused persons, to ensure that such accused have an equal opportunity to participate in the legal consequences that may follow upon them being declared unfit to plead and stand trial. Protection of the law furthermore requires tailored support for accused persons having been found unfit to plead and stand trial; support that ultimately ensures access to justice by focussing on the ability-specific needs of the individual involved.

At present, and as was succinctly illustrated in the above exposition, the provisions related to an accused person presumed unfit to plead and stand trial in Hong Kong, are scattered over various Ordinances: the Criminal Procedure Ordinance Cap 221, Magistrates Ordinance Cap 227, Mental Health Ordinance Cap 136, Juvenile Offenders Ordinance Cap 226 and Prisons Ordinance Cap 234.[110] Particularly onerous is the interface between the provisions of these Ordinances and especially with regard to the various orders a court can make and how and by whom these orders ought to be effected, renewed, discharged, or appealed/reviewed.[111] Moreover, and as was evident from the discussion in the preceding part, while comprehensive provision is made for the relevant procedures in the Court of First Instance, sparse guidance exist for instances where an accused suffering from a mental disability or intellectual impairment is unfit to plead

[108] ibid.

[109] ibid [10], [14].

[110] Only the most important provisions were considered in this article.

[111] See, for example, section 76 of the Criminal Procedure Ordinance Cap 221, read together with Parts IIIA, IIIB and IVA of the Mental Health Ordinance Cap 136.

and stand trial in a Magistrate's Court, District Court, and Juvenile Court. It is therefore submitted that the current legislative framework for determining whether an accused person is unfit to plead and stand trial, do not promote accessibility and transparency to the end of empowering such an accused person to participate in the legal proceedings and subsequent decisions. In noting a similar shortcoming in the comparable laws and provisions currently enacted in England and Wales, the Law Commission stated that for all those affected by a finding that an accused is unfit to plead and stand trial, including the complainants, witnesses and family members:

> The complexity and inaccessibility of the current law is a significant barrier to their engagement and undermines their confidence in the criminal justice system. Complainants and those who support them, who are themselves often volunteers with no legal expertise, would welcome an easy to locate, readily understandable test, set out in statute. Leaving the test to be hunted out in case law is inconsistent with efforts to make the criminal law accessible to those affected by it.[112]

However, the best and most appropriate way in which to support and provide for accused persons suffering from a mental, intellectual, or cognitive disability or impairment at trial, and specifically in ensuring their equal participation before and protection of the law, remain vexed. A fine balance must ultimately be achieved between the interests of society and the rights and welfare of an accused person rendered vulnerable by mental disability or intellectual impairment. This problem was described as follows by an unknown author in a 1967 note published in the *Harvard Law Review*:

> The problems of incompetency law reflect a basic ambivalence in society's attitude toward mentally ill criminal offenders. On the one hand, there is a feeling that persons who are mentally disabled ought to be sheltered from the severity of the criminal process while, on the other, there is a desire to protect society by confining and punishing persons thought guilty of criminal conduct.[113]

[112] Law Commission, *Unfitness to Plead Volume 1* (Law Com No 364, 2016) [3.27].
[113] Note, 'Incompetency To Stand Trial' (1967) 81:2 Harvard Law Review 454, 472.

A. Intermediaries

One way in which to address the obstacles created by the current Hong Kong legislative patchwork of ordinances and provisions governing the proceedings, determinations, and legal consequences involving accused persons unfit to plead and stand trial, is to make provision for the appointment of intermediaries. An intermediary can be defined as "a communication expert whose role is to facilitate a witness's or defendant's understanding of, and communication with, the court".[114] To date, Hong Kong law does not make provision for the appointment of intermediaries in any of its legal proceedings involving vulnerable categories of persons. This is regrettable as a statutory entitlement to the assistance from an intermediary can be particularly helpful to those accused suffering from a mental, intellectual and/or cognitive disability, and who is otherwise not necessarily also unfit to plead and stand trial. For these accused, a statutory entitlement to receive assistance from an intermediary may facilitate a more accessible proceeding in which such accused persons can exercise their autonomy and effectively participate and vindicate their fair trial rights.[115] In fact, the appointment of intermediaries for accused persons suffering from a mental, cognitive or intellectual disability was one of the recommendations made by the English Law Commission in its 2016 report,[116] and it is also consonant with the spirit of the United Nations Convention on the Rights of Persons with Disabilities.

For example, article 12(2) of the Convention requires State Parties to "recognise that persons with disabilities enjoy legal capacity on an equal basis with others in all aspects of their life".[117] Legal capacity is a distinct concept from mental capacity and "involves the capacity to be a holder of legal rights, and the full protection of those legal rights in the eyes of the law".[118] Given this broadening perspective on the effective and meaningful exercise of human capacities and abilities before the law, the focus can no longer be on the threshold determination of whether an accused person is fit to plead and stand trial, but should rather focus on "identifying and implementing the supports necessary to help an individual [to]

[114] Law Commission, *Unfitness to Plead Summary* (Law Com No 364, 2016) [1.27].

[115] ibid; Law Commission, *Unfitness to Plead Volume 1* (Law Com No 364, 2016) [2.31]-[2.94].

[116] Law Commission, *Unfitness to Plead Summary* (Law Com No 364, 2016) [1.28], [2.31]-[2.94].

[117] Convention on the Rights of Persons with Disabilities (adopted 13 December 2006, entered into force 3 May 2008) 2515 UNTS (CRPD) art 12(2).

[118] Anna Arstein-Kerslake, Piers Gooding, Louis Andrews, and Bernadette McSherry 'Human Rights and Unfitness to Plead: The Demands of the Convention on the Rights of Persons with Disabilities' (2017) 17 Human Rights Law Review 399, 405.

exercise that absolute right to legal capacity".[119] Intermediaries can certainly play an important supportive role to accused persons suffering from a mental, cognitive or intellectual disability, and who find themselves entangled in the labyrinth of Hong Kong's current legislative framework for fitness to plead and stand trial.

B. Diversion from the criminal process

Where an accused person has been declared unfit to plead and stand trial, however, the support of an intermediary may not be sufficient. Moreover, the continuation of the legal process for the purpose of making a factual finding on whether that accused person had committed the act alleged, may in itself be unjust as it amounts to a differential process, stripped from some of the procedural safeguards that underpin the adversarial criminal trial.[120] For example, in the exposition of the Hong Kong legal framework above, and with reference to the recent English case of *R v Orr*,[121] it was highlighted that although a factual finding on whether an accused person did the illegal act alleged, may hold the advantage of providing closure to the criminal proceeding, it remains problematic as it constitutes a legal pronouncement against which the accused person is not able to properly defend him or herself, and upon which legal consequences may nonetheless follow.

Moreover, in the context of Hong Kong, an agenda for the active prosecution of 'mentally ill' persons is seemingly at the order of the day. For example, article 5.10 of the Hong Kong Prosecution Code (2013) provides as follows:

> The criminal justice system operates to protect both the community and individual members of it. From time to time the prosecution may consider it appropriate to charge mentally ill persons with applicable offences principally in order to invoke the court's jurisdiction to make beneficial orders for the management of the mentally ill, their protection and the protection of the community.[122]

Thus, at stake here is the potential benefit of the court's jurisdiction to make determinations and impose appropriate orders in terms of the provisions of the Criminal Procedure Ordinance Cap 221 to ensure that the interests of society are

[119] Arstein-Kerslake, Gooding, Andrews, and McSherry (ibid) 406; Convention on the Rights of Persons with Disabilities (adopted 13 December 2006, entered into force 3 May 2008) 2515 UNTS (CRPD) art 13(1).

[120] Arstein-Kerslake, Gooding, Andrews, and McSherry (ibid) 404.

[121] *R v Orr* [2016] 4 WLR 132.

[122] Hong Kong Prosecution Division, Hong Kong Prosecution Code (2013) article 5.10.

protected from possible harm or violence, and that accused persons suffering from a mental disability receive the necessary treatment.

A case in point is *R v Chan Ming Kwok* in which the applicant had a long history of mental illness and was charged with one count of attempted murder and two of wounding with intent. The trial judge, in making a hospital order under section 45 of the Mental Health Ordinance Cap 136 stated:

> I find it impossible to forecast today that it would be safe to release the defendant on any particular date. His condition can be controlled by regular medication. Whilst he is detained the administration of the appropriate drugs can be ensured. Once he is released, if he ceases to take the necessary drugs he may be expected to relapse, in which case as his previous conduct shows he can be highly dangerous. I know of no order I can make which will ensure that he takes the drugs once he is released. In the circumstances I refrain from specifying the period during which the defendant is to be detained. I leave it to the medical authorities, once they are satisfied that it will be safe to release the defendant, to take the necessary steps under the Mental Health Ordinance for his discharge.[123]

The majority of the Hong Kong Court of Appeal agreed with this finding and emphasised that the applicant was a danger to the community if he was not under medication and proper treatment.[124]

Yet, the potentially dire consequences such a prosecution policy may have for a vulnerable accused should also be noted. The appellant in *HKSAR v Meijne, Camilo Arturo*) had a long history of bipolar affective disorder and experienced difficulty in coping with stressful situations. In the lead-up to the incident resulting in a charge of indecent assault, which was later amended to a charge of common assault to which the appellant pleaded guilty, the appellant displayed erratic behaviour and obvious signs of an escalation of his mania.[125] In fact, the appellant's treating psychiatrist recommended that he be admitted to a psychiatric hospital but the appellant went missing for three days and was not seen until the day of the incident. The incident was captured on CCTV footage and involved the appellant, standing inside a lift, allegedly 'pushing' the victim's right breast once

[123] *R v Chan Ming Kwok* [1987] HKCA 194; [1987] 3 HKC 222; CACC 82/1987 (11 August 1987) [2].
[124] ibid [3].
[125] *HKSAR v Meijne, Camilo Arturo* [2015] HKCFI 131; HCMA 274/2012 (30 January 2015) [1]-[2].

with his left hand before pressing the button to close the lift door.[126] At trial, the appellant was found fit to plead and stand trial but an order was made, in virtue of the ongoing concern for his mental health, that he be remanded in custody for observation in a mental health hospital and that the findings of two psychiatrists be brought before the court. The psychiatric reports confirmed the appellant's mental health condition, which had by that time deteriorated, and it was suggested "in the event of a conviction, [that] a hospital order for a period of three months should be imposed".[127] The two psychiatrists nonetheless declared the appellant mentally fit to plead.[128] The appellant was subsequently convicted and sentenced to twenty-one days' imprisonment reduced to fourteen days for the plea of guilty. A suspended sentence of one-month imprisonment for an assault occasioning actual bodily harm was also triggered. The appellant served his entire sentence at the Siu Lam Psychiatric Centre, after which he was referred to the Pamela Youde Nethersol Eastern Hospital where he was detained for a further sixty-two days as an involuntary patient.[129]

On appeal, Judge Zervos for the Hong Kong Court of First Instance held that the appellant's psychiatric history was relevant to the question whether the appellant should have been prosecuted, the alleged incident of the assault, and also the determination of an appropriate and just sentence.[130] On the first ground, Judge Zervos emphasised that the magistrate who hears a fitness to plead application should also be the magistrate who ultimately presides over the trial. This was not so in this case, as the Principal Magistrate had called for the psychiatric reports to determine whether the appellant was fit to stand trial and another magistrate was thereafter assigned to preside over the trial. This was described as 'regrettable', as the two psychiatric reports made it clear that while the appellant was fit to stand trial at that moment, his mental health condition fluctuated and he was subject to manic episodes and relapses.[131] This information was therefore relevant not only on the question whether the appellant was fit to stand trial at arraignment, but more generally in terms of his mental condition during the course of the trial. With regard to the second ground it was found, upon a close scrutiny and analysis of the CCTV footage, that the appellant did not in fact push the complainant as alleged, but that he rather held out his left-hand gesturing to the complainant not to enter the lift, and that his outstretched arm and hand made contact with the

[126] ibid [2].
[127] ibid [6].
[128] ibid [6].
[129] ibid [3].
[130] ibid [8].
[131] ibid [13].

complainant above her right breast for a mere second.[132] Of this Judge Zervos stated:

> As far as I am concerned it was not a push and there was no aggressive action by the appellant with his left hand. It was simply a gesture not to come in with his left hand reached out. It was also clear that it was the slightest of contacts which occurred in less than a split-second. Whilst I can understand the woman being offended by not being let in the lift, there was no physical action by the appellant who probably, because of his mental state, wanted to be left alone in the lift.[133]

Thus, given the appellant's mental condition and that the alleged incident involved minimal contact, Judge Zervos concluded that the facts of the matter did not support a criminal conviction and the conviction was quashed and the sentence and order to activate the suspended sentence was set aside.[134]

These two cases illustrate the practical difficulty of having a prosecution policy that seemingly encourages the prosecution of those who may suffer from mental, intellectual or other cognitive disabilities. While it may sometimes be effective to use a criminal prosecution to bring somebody under the protection of the court, it must also be remembered that the positive factual finding for the alleged act or omission performed by an accused subsequent to a finding that the accused is not fit to plead and stand trial, has consequences. These consequences may be dire for those who are not a danger to society and who would not benefit from becoming entangled in a labyrinth of Ordinances and role-players who make life-decisions on their behalf. It is for this reason that the English Law Commission recommended that a court should, subsequent the finding that an accused person is not fit to plead and stand trial, have the option *not* to embark on the deliberation as to whether the accused committed the act or omission as charged.[135]

In other words, to divert from the matter from the criminal justice system. This would not only be consonant the fundamental principle that an accused be present at his or her own trial and that the accused effectively and meaningfully participate in his or her own defence, but it would also be respectful and accommodating of those who are not able to fully and meaningfully participate in a criminal proceeding because of mental, intellectual, or cognitive incapacities, and irrespective of whether the incapacity is of a temporary or a permanent nature.

[132] ibid [22].
[133] ibid [22].
[134] ibid [22]-[23].
[135] Law Commission, *Unfitness to Plead Summary* (Law Com No 364, 2016) [1.18]-[1.19].

Where a criminal process for the determination of factual guilt or innocence nonetheless continue after the accused person was found to be unfit to plead and stand trial, it was already emphasised above that every effort must then be made "to afford a defendant whose capacity may be in doubt such adjustments to the proceedings as he or she reasonably requires to be able to participate in the full criminal process, and to maintain that capacity for the whole of the process".[136]

The English Law Commission further recommended that courts exercise a judicial discretion not to proceed with a hearing to consider the allegation following a finding that the accused is unfit to plead and stand trial, and explained that such a discretion:

> [S]hould be subject to an interests of justice test, to be applied by the judge taking into account various factors, including: (1) the seriousness of the offence; (2) the effect of such an order on those affected by the offence; (3) the arrangements made (if any) to reduce any risk that the individual might commit an offence in future, and to support the individual in the community; and (4) the views of the defence and the prosecution in relation to the making of such an order.[137]

Yet, it was also submitted that the exercise of such a judicial discretion should not prevent the prosecution from applying for leave to resume prosecution, in appropriate cases and where the accused subsequently regains capacity for trial.[138]

Moreover, and as was indicated before, it is particularly problematic for a factual finding to be made as to whether an accused had committed the act or omission charged, if only the *actus reus* of that act or omission is considered. The English Law Commission stated in this regard that "the unfit individual is substantially disadvantaged in comparison to a defendant facing the same allegation in full trial", as the ability of that unfit individual to rely on common defences such as self-defence, accident or mistake will be significantly restricted.[139] The Commission subsequently recommended that where the factual determination

[136] ibid [1.15].
[137] ibid [1.71].
[138] ibid [1.72].
[139] ibid [1.20].

is pursued, that the prosecution be required to prove all elements of the offence beyond a reasonable doubt. This, it was submitted,

> [W]ould afford individuals who lack capacity the same opportunity to be acquitted as is enjoyed by defendants who have capacity, enabling them to engage all available full defences…The resulting finding at the hearing would not be a conviction, since the individual who lacks capacity is unable to participate effectively in trial, but an alternative finding that the allegation is proved against him or her.[140]

V. Conclusion

It is clear from the critical and legal-historical analysis in this article that the Hong Kong legal framework for 'fitness to plead and stand trial' has not been the subject of critical revision and reform to bring the relevant substantive and procedural law in line with contemporary understandings and awareness of mental health and cognition. This, it was noted, affects the fair trial rights of vulnerable accused persons suffering from a mental, intellectual, or cognitive disability.

The shortcomings identified in this article with regard to the current Hong Kong legislative framework for fitness to plead and stand trial, included substantive matters like the legal test for determining whether accused persons are able to effectively and meaningfully participate in the legal proceedings against them, procedural matters relating to the process for this determination and the legal consequences that may ensue, as well as the elaborate structure of ordinances and policy on which this Hong Kong legislative framework is based. It was noted, for example, that Hong Kong law excessively pathologises human (in) capabilities, which not only contributes to the stigma surrounding mental illness but may also result in a distorted view of the vast continuum on which a person's capacity and incapacity, or ability and inability to make decisions, can exist. It was also shown that the determination of whether an accused person can effectively and meaningfully participate in the criminal proceedings against him or her, has become a sufficiently complex matter that can no longer be left to the jury. Particularly problematic in Hong Kong is that differing procedures exist in this regard for trials at Magistrate's Court and District Court level, and for proceedings before the Court of First Instance. Such differing procedures are unnecessary, breed inconsistency, and are particularly time-consuming in the Court of First Instance where a jury is tasked to make the determination. And finally, with regard to the consequences that may follow upon a determination of unfit to plead and

[140] ibid [1.73]; Law Commission, *Unfitness to Plead Volume 1* (Law Com No 364, 2016) Chapter Five.

stand trial, it was evident that much more can be done to ensure the effective and equal participation of vulnerable accused before the law.

This evaluation of the Hong Kong legislative framework was informed by the comparable legal developments in the laws of England and Wales, as well as the provisions of the United Nations Convention on the Rights of Persons with Disabilities. The recommendations subsequently made essentially require of the Hong Kong legislative framework to better reflect contemporary comparative and international understandings and awareness of mental health and cognition, and to ensure equal recognition before the law of all persons, including persons suffering from mental disabilities or intellectual impairments. Such legal reform and development will not only align Hong Kong law with the laws of England and Wales—which are important for historical reasons and for the coherent and systematic development of the laws of Hong Kong—but will also ensure that Hong Kong meets its obligations under the United Nations Convention on the Rights of Persons with Disabilities.

It is submitted that the fair trial rights of vulnerable accused persons suffering from a mental, intellectual, or cognitive disability, should be a critical priority on the agenda of Hong Kong legislators. Yet, the plight of such accused persons should not be made exceptional; that is a differentiated jurisprudence that stands distinct from mainstream fair trial rights. This is because *fairness*, in adversarial systems like that of Hong Kong, also requires *equality* before the law, as well as equal protection of the law.[141] Our longstanding commitment to fair trial rights, must therefore evolve with contemporary understandings of mental health and cognition to ensure equal recognition before and equal participation in the law, for all persons, including those suffering from mental disabilities or intellectual impairments.

[141] Michaël van der Wolf, Hjalmar van Marle, Paul Mevis and Ronald Roesch, 'Understanding and Evaluating Contrasting Unfitness to Stand Trial Practices: A Comparison between Canada and the Netherlands' (2010) 9:3 International Journal of Forensic Mental Health 245, 249.

A Concept of Personal Autonomy Fit for Contract Law

*Diogo Tapada dos Santos**

Abstract

The interplays between autonomy and many areas of law are somehow evident but many times redundant and unclear. In this paper, I offer an account of personal autonomy that can be useful in reading private law phenomena, especially focusing on the doctrine of contract as promise. This doctrine, which assimilates contracts and promises, poses two challenges to the ideal of personal autonomy: first, how can autonomy justify or require the ability to be bound by promise or contract; second, how can the possibility to change one's mind, which is a virtue of the autonomous life, be reconciled with the bond created by contracts. By defending that personal autonomy is an ideal of self-authorship, and that the autonomous person authors her own life, being emotionally and intellectually capable of committing to a sufficient variety of long- and short-term choices, which she will consider without uncalled external interferences nor impositions and with respect for a condition of integrity, I will address both concerns, aiming at proving that a thick ideal of autonomy is capable of justifying the practice of contracting and is not in opposition with the strong bond created by contracts. The

* Diogo Tapada dos Santos, MJur (Oxon). Junior Associate at Linklaters LLP (Lisbon) and Guest Lecturer in Civil Procedure (University of Lisbon School of Law). This paper corresponds to the dissertation presented for the degree of Magister Juris (Oxford, 2015), supervised by Professor Leslie Green. The author is thankful to Clara Martins Pereira for her contribution in discussing and revising the dissertation.

debate will primarily focus on promises, and thereafter its arguments are applied to contracts' analysis.

I. Why Autonomy in Contract Law?

The idea of personal autonomy has long occupied a prominent place in philosophy and is certainly implied in many aspects of legal theory. However, it is many times used in an ambiguous and circular way: people seem satisfied to say that a person is autonomous when a person has autonomy, without elaborating on what that autonomy would mean or imply, and that confers a high degree of elasticity to the concept.[1]

When the debate is moved to the field of law, one soon finds out that autonomy-based accounts of the law tend to be simplistic. Most lawyers are confronted with the very general idea that, because the person is autonomous, the law seeks to create room for the exercise of her personhood. This is a poor account that does not tell us much and does not commit to any vision of autonomy in particular.

An idea of autonomy, even if slim or thin, can be said to exist in all fields of law (again, within the simplistic framework). Public law has rules that limit the state scope of intervention, securing negative liberty. Criminal law works as a coercion mechanism to avoid interferences between persons that diminish the autonomy of ones' life [2]—perhaps the clearest example is the criminalization of kidnapping of persons, a crime that in practice limits someone's practical autonomy.[3] Diversely, private law is not merely a mechanism of negative enforcement of some autonomy, but is more essentially a mechanism of positive enforcement of autonomous actions. This can be seen in aspects such as the acceptance of the trust as a mechanism to transfer wealth, but is more evident in contracts. However, there is not, I believe, a special relationship between autonomy and private law. What is worth pointing out is that, as happens with public law or any kind of law for this effect, there are aspects of the legal regime that both seek justification in and struggle with the ideal of autonomy. Thus, each particular legal institution may have a relevant interplay with the ideal of autonomy.

Brevitatis causae, I do not intend to offer a full account of the intersections of autonomy and private law, elaborating on some sort of catalogue of autonomy-

[1] For a summary on the broadness of uses associated with autonomy, see Gerald Dworkin, The Theory and Practice of Autonomy (Cambridge University Press 2001) 6.

[2] I am assuming that criminal law works primarily as a mechanism of general prevention by exerting a threat that is sufficient to dissuade criminal activity.

[3] The question of whether or not someone who is physically constrained remains morally autonomous is left open.

driven or autonomy-limited institutions, nor discuss whether autonomy is the driving principle of the institution of contract.

Coming to the delimitation of the scope of this paper, I expect to arrive at an account of personal autonomy, a concept that helps to understand the idea of 'contract as promise'. This is not new: Kimel undertook a similar exercise[4]. Nevertheless, I hope to bring some elements to the debate, which relevance lays in the challenges that the contractual promissory events pose to the ideal of autonomy. The brevity of this paper requires that only one problem is addressed. Nevertheless, the theory of contract as promise is far from being uncontested and there is more to a contract than its promissory facet, or more to a promise than its contractual potential. Therefore, some assumptions are made for the sake of the argument with a view to enable an elucidating debate on the idea of personal autonomy that is working as a justifying factor of private law institutions.

Likewise, my account of personal autonomy will not certainly be an all-encompassing explanation, considering what has been said before—it is oriented at analysing a specific phenomenon. The value of this exercise relies, in my view, on the fact that contractual obligations are a good illustration of the interplays between a private law institution and autonomy: one has to be able to explain if and why does an autonomous person require the ability to be bound by contract, while solving the paradox that an autonomous person bound by contract will not be able to easily change her mind. By doing so, we are elucidating by example a potentially broader construct about the role of autonomy in private law institutions and we can establish a thick concept of autonomy that sustains the binding force of contracts.

An example may be useful. *A enters into a lease contract with B for the period of three years. After one year, A has a change of mind*[5] *and B insists A has to keep the lease two more years*. A simple view of autonomy will struggle with such case: A was certainly autonomous when entering into the lease, and as such the law has protected A's interest in this contract by making it binding to B, who had agreed to lease to A. However, as an autonomous being, A can have a change of mind; why does not the law allow for the unmotivated rescission of the lease, protecting B against A's change of mind? A thick concept of autonomy should be able to explain this simple case without compromising any relevant moral value of contracting and of living the autonomous life.

[4] Dori Kimel, 'Promise, Contract, Personal Autonomy, and the Freedom to Change One's Mind' (2013) Oxford Legal Studies Research Paper No. 19/2013.

[5] Note that this change of mind has to be legally irrelevant: it cannot be, for instance, a change in circumstances that legally allows for the rescission of the lease.

II. Contract as Promise: Introductory Remarks

Some notes on why contract as promise is of relevance and asks for an autonomy-based account of law are required. The background conception is that a contract is frequently seen as an interchange of promises[6] (hence the nomenclature of promisor and promisee, which illustrates more than anything this understanding), which is cause for concern because contracts in such sense are an exercise of self-authoring that may be subject to a change of mind. From that, one easily questions how a person can be bound by contract, when a contract is an autonomous promise subject to the possibility of a change of mind.

Let us lay down the problem. The first idea is that a contract is a legally acknowledged promise. The second is that a promise has moral worth by virtue of it being an exercise of autonomy. There is an assumption that self-authorship is an essential manifestation of the idea of autonomy, and that promising is a prominent way of self-authoring one's life. This second idea requires discussion of two aspects I will analyse on Chapters III and V: first, the fact that self-authorship is part of the ideal of autonomy, and second, the statement that promising has anything to do with self-authorship.

The third idea is that, for an autonomous person, the ability to commit oneself is as worthy as the ability to change one's mind. It is a legitimate claim that not only committing oneself, on the assumption that promising favours autonomy, is good, but also that the possibility to change one's mind is a requirement for an autonomous life. How can two good things annul each other? For promising supposedly presupposes that a change of mind will not be allowed an impact on a commitment, and this opens up a paradox: the more a person promises, the more a person is autonomous on one side, and the less a person is autonomous on the other; because the more a person promises, the less a person can act on a change of mind.[7] This apparently invalidates future alterations to the contract, starting by preventing the promisor to change a promise and even by trying to avoid a failure to perform.

In what regards the first idea, I do not intend to give a full account of the conception of contract as promise, as that would exceed the scope of this paper. I will say that it is almost factual that a contract is an exchange of promises: a promisee

[6] It is worth pointing out that the statement that a contract is an interchange of promises does not automatically endorse the liberal theory of contract. It happens that case law and many legislative texts address the parties of the contract as promisor and promisee, and label both the contractual obligation and the object of the contractual performance as a promise.

[7] N.B. I say "act on a change of mind" in order to avoid for now the debate on whether a change of mind, at all, is possible. It seems that a promise will not impede a change of mind, unless a person is of such a moral character that a complete adhesion of spirit to a promise crystallises the mind for all effects.

seeks a promisor, who will undertake an obligation that is legally enforceable. The main difference between contract and promise is precisely in the enforcement mechanisms available to the promisee, for to a legally irrelevant promise there are no legal enforcement mechanisms available. Thus, contracts can be seen as a subspecies of promises that the law vests with certain peculiarities in regime. At first it seems that the promises qualified as contracts are not different from promises in general: both are relationships by which one person is obligated to fulfil the undertaken in the benefit of another person. However, contracts are better seen as qualified promises, to which the law and the courts apply specific tests as to give them legal relevance. The availability of mechanisms of legal enforcement is a mere consequence of the legal relevance of a promise, and therefore they are insufficient to argue that contract and promise are two different realities. The legal relevance of a promise, i.e., its qualification as a contract, is established with recourse to tests that normally rely on formalities (as is the case of promises under deed) or on factors that indicate a relationship worthy of legal protection (as the common law doctrine of consideration). Although these may qualify a promise, there is no significant change in the basic nature of the practice. In fact, the mechanisms of legal enforcement of promises (namely, specific performance and damages) work as to coerce the promisor to stick to his promise, either by giving him a direct order to perform or by making it disadvantageous for him not to perform, to the point that performance comes as a lesser evil in his mind. Likewise, promises outside of contract can be enforced by recourse to social pressure or by appealing to the promisor's character or morality. Therefore, one can assume for now that a contract is indeed no more than a species of promise, subject to a particular regime as regards its formation and fulfilment.

III. A Concept of Autonomy

A. The need for a concept

It is important to start by making clear what the use of the word autonomy conveys. Without this clarification, one can easily slip into simplistic stances such as 'a person can enter a contract because it is up to her, by virtue of her autonomy, to stipulate legal relationships'. What is then the idea of personal autonomy that can be relevant to explain the legal phenomenon of 'contract as promise' here at stake? I wish to delimit this incursion on a concept of autonomy by functionalizing the debate, i.e., by accepting that the definition of any concept must take into consideration the prudential uses it is going to serve. Naturally, being teleologically oriented does not imply a deviation from the bigger picture, i.e., the concept of

autonomy in a "pure" form (if such a form is possible, probably it can only be validated in very abstract terms).

As regards the issue of 'contract as promise', the autonomistic conception has to be able to explain two distinct aspects, as already noted. Firstly, the concern that an autonomous person may, by binding force of a contract that is no more than a promise comprising certain rules for action or abstention, limit her concrete capacity for autonomy in certain cases. In our case, this first concern is reflected in B committing her property to a three-year lease, excluding all other uses a property owner has available. Secondly, the perplexity that an autonomous agent can be so bound by contract that his contractual promises can be held against him, even against his own will. In our case, this is illustrated by A's change of mind one year into the contract. The first aspect relates to the omnipresent question of whether the autonomous person can, by her own will, become less autonomous—or, looking at it from an alternative perspective, whether the assumption of compromises (such as promises) is in fact the attitude that best conforms with the autonomous life. The second aspect regards the possibility for a change of mind and its apparent lack of influence to legal life.

B. Concerns and constraints when defining a concept of autonomy

In order to be able to address these aspects, any conception of autonomy should satisfy some concerns regarding its plausibility and coherence. [8]

One must be aware that the concept sought is going to deal with the legal system (taken as an abstraction here, for I am not considering any particular legal system). There is a certain degree of criticism to the legal rules and their system allowed, especially in the event one finds that the rules under analyses are autonomy-impairing. Notwithstanding, there may be a presumption that the legal rules are a product that reflects the best regulatory solution found to date, and in so far as possible, they should be reconcilable with the concept of autonomy.

Furthermore, Dworkin notes that a concept of autonomy must not be such that it 'makes it impossible or extremely unlikely that anybody ever has been, or could be, autonomous'.[9] Thus, autonomy—even if an ideal—has to be seen in realistic terms. This is perhaps the most relevant concern, being a deal breaker to the whole question. I wish to extend this concern further: autonomy has to be able to frame the phenomenon of promising, given that promises are part of the everyday life and are in themselves an exercise of the autonomous life. It is impossible to

[8] These are inspired by Dworkin (n 1) 7–9.
[9] ibid.

try to explain promising (and contracts as promises) in an autonomy-inspired framework if one is to argue, further on, that the very idea of obliging oneself by a promise harms autonomy. Such an endeavour would lack logical coherence unless one finds a good justification to say that an anti-autonomous practice is widely allowed in societies and jurisdictions that praise the value of autonomy. For instance, everyone will agree that B's right to lease her property exists, and that it is a good thing. Alongside this idea, the concept of autonomy must have 'normative relevance', i.e., it must make clear its usability as a philosophical device.[10]

Besides, any concept of autonomy must be able to explain why people would consider being autonomous good and has to be reconcilable with other goods (competing values). I am here following Dworkin's refusal of the strongest constraint in the value conditions, that if accepted would require prove that autonomy is the supreme good. The reason for this option can be disclosed now, in spite of it being necessarily developed further on: autonomy can hardly be the supreme good, considering that some competing goods may have equal worth, e.g., the reliance of the promisee on a promise can be opposed to a change of mind of the promisor. If we recall the lease, the fact that B has leased her property is good but not the only relevant good for law; in fact, governments sometimes limit or put constraints on the right to lease because there are competing goods (e.g. market stability or house availability) and B's autonomy is sometimes not enough to trump other competing values.

As a weak constraint to the concept, Dworkin proposes that it should respect some ideological neutrality, meaning that the concept must be valuable for various ideological standings. Although I see the necessity for such a constraint, in the sense that a concept entrenched in ideology would not be a satisfying justification for many bystanders, it should be taken carefully. A concept that is too broad and ideologically uncompromised can end up not satisfying anyone's system of beliefs. Dworkin seems to be sensible to these concerns and that is why he opts for a weak constraint, allowing for differences between ideologies regarding particular aspects of autonomy to be relevant, such as its absolute or relative value.[11] For instance, in a country with controlled rent laws, the landlord can still be considered autonomous in leasing his property. Finally, the concept of autonomy has to some extent to respect the set of judgements one has about the idea of autonomy, which in our particular case implies regard to assumptions such as 'contracts without consent

[10] Dworkin (n 1) 8–9.
[11] ibid 8.

are invalid because there was no autonomy' or 'all legal restrictions observed, the parties can autonomously establish the contractual provisions they wish'.

C. At last, a concept of autonomy

After the previous remarks, one can adventure into offering a concept of autonomy with which the idea of contract as promise can be explained and justified. There is little room for originality here: many authors have advanced definitions and ideas, and my goal is to have a working concept that serves the main purposes of this paper.

The fundamental idea that seems to permeate all conceptions of autonomy draws from the etymology of the word: autonomy relates to a norm (*nomos*) that is self-imposed (*autos*).[12] In simple terms, the autonomous person is the one who sets her own rules. This is rather vague and leaves unanswered what is the nature of these rules. Besides, it assumes autonomy is an un-empirical deliberation mechanism, which is action-independent. Perhaps this confusion is related with the appropriateness of the concept of autonomy that, one must recall, first started in political philosophy as an attribute of Greek cities, then of collective entities such as the state, and finally of individuals. As in all conceptual extensions, *mutatis mutandis* is key: for instance, comparing an individual to a legislator of his own life is impacting, but still a metaphor. Thus, the concept should grasp something more than an idea of setting rules for oneself, rejecting that the autonomous life is merely an experience of defining a set of rules and living by them.

I believe it is more precise to claim that the autonomous person is one who authors her own life, be it by means of self-created rules, adhesion to a life plan, or even a conscientious decision not to follow any rule. Still, the element of self-authoring through auto-normativity is fundamental, especially if one accepts that to set norms to oneself is to "weave the tapestry" of one's own life. There are competing conceptions of autonomy, but their focus is on precisions about the sense of self-authoring one's life more than on denying that autonomy is about self-authoring.

The first aspect in delineating a concept of autonomy is to engage in the debate between procedural and normative accounts of autonomy. A procedural view will argue that a person is autonomous when she acts within a procedure (a set of formal conditions) that has in view an autonomous action, i.e., the emphasis of this account is on the exercise, the way by which the person reaches a decision. On the other hand, a normative perspective will argue that it is not right or enough to focus on the form of an exercise or decision, but it is necessary to observe

[12] Such is the account of Dworkin (n 1) 12.

substantive conditions to the decision so that it can be deemed autonomous; that is to say, autonomy exists when the decision abides to a certain norm that is reflected in substantive conditions. A proceduralist will be satisfied when a person decides following a procedure that assures an autonomous outcome. If we think about contract law, a proceduralist will say A and B entered into their agreement in an autonomous way if they followed the legal formalities. A normativist will further ask the decision to be assessed according to a set of conditions. In our case, besides any formalities, A and B are to be under no duress or mistake. The normativist accounts vary on the conditions that are required. Moreover, it is possible to be a normativist and still to acknowledge the importance of the procedure under which the exercise of autonomy occurs. In my view, to judge a person to be autonomous one has to look at the procedure, and at the content of the decision and its agreement with the substantive conditions for autonomy. One without the other offers an incomplete picture of the idea of self-authorship: to claim that someone is authoring their own life requires the author to be autonomous and his agency to respect autonomy. If the author is not autonomous, then he is not in fact the author. And if his agency does not respect autonomy, then his decisions are not worth of his authorship—rather, they are an emulation of an autonomous life. Later I will try to show how two substantive conditions of the autonomous life—responsibility and integrity—are necessary to identify someone as autonomous.

A second aspect to take into consideration is the debate about the distinction between first-order and second-order desires or preferences. Roughly, the first-order preferences are those that are manifested immediately, ruling concrete exercises of one's autonomy, and second-order preferences belong to the realm of self-reflection, translating a certain sense of identity or of long-term options. Dworkin relies on such a distinction to deny that autonomy only concerns second-order reflection.[13] It is true that an individual builds desire upon desire, that these are interrelated and that, as such, one could devise a "chain of desires". His example is of the man who smokes, desiring so on a first-order basis, but wishing he did not on a second-order reflection. Nevertheless, he recognises the criticism to which the idea of second-order preferences can fall victim, mainly that first-order desires do not need considerations about second-order reflection and that recognising a second level of preferences allows for a never-ending enquiry into the mind of the person, who will always find new levels of reflection.[14] A third obstacle to this distinction seems to be its artificiality. Without entering into the science of cognition and volition, how can we establish with clarity levels of preferences? By questioning the person? Or by adopting an analytical posture and trying to

[13] Dworkin (n 1) 18.
[14] ibid 19.

make sense of an individual's actions and motives? It all appears odd and it is. When it comes to preferences or self-reflection, I believe it is impossible to precisely individuate such first- and second-order levels. It is safer to regard an action or omission, in order to assess the autonomy of the agent, in its entirety, considering it to ontologically carry the whole of self-reflection. Autonomy then calls for a holistic reading of the design of the person's life, pondering lingering tendencies, emerging inclinations, long-term reflections, non-negotiable principles of conduct and short-term options. This implies that the judgement of whether a person authors her own life should take into account the whole of the life and the whole of the person, and cannot be centred in a casuistic analysis of actions and omissions.

D. Conditions for autonomy

So far, we have seen that the autonomous person is one who is capable of authoring her own life, and that both the process and the outcome of every "activity" (for the lack of a better word) of self-authoring is relevant to deem the person autonomous in a holistic account of her life. These reflections presuppose a background state of affairs for autonomy: the capacity for it or the conditions of autonomy, as Raz labelled them.[15] Raz says that the person, to be autonomous, must have appropriate mental abilities, an adequate range of options to choose from and be independent. I accept these conditions, which will be analysed and developed, and will further add integrity as a fourth condition.

The condition of having appropriate mental abilities calls for a judgement on the intellectual and emotional capacities of the person to determine if she has 'the mental abilities to form intentions of a sufficiently complex kind, and plan their execution'.[16] Contract law fully embodies this condition by only enforcing promises of agents that have legal capacity.

As to having an adequate range of options to choose from, it can be translated into the necessity of having a diversity of plans of life at the individual's hand. The autonomous person has to be able to choose what sort of life she will lead and, apart from natural constraints (although even those can be circumvented at times), all possibilities are supposed to be available. A person cannot be said to be the author of her life if there is no room for choice between alternatives, or better, if the whole of the life's plan comes from a source outside the person's volition. However, practicability recalls us that a requirement for all possibilities to be available is too strong: while in theory it is possible to see as desirable a full world of potential for self-authorship, in practice it seems evident that not every possibility can be available. This is due to various reasons: firstly, some possibilities

[15] Joseph Raz, *The Morality of Freedom* (Oxford University Press 2003) 372–373.
[16] ibid 372.

are mutually excluding, meaning that by choosing a certain norm the person is also choosing not to choose a norm incompatible with the former; secondly, not only nature sets some restraints (v.g., I could choose to run a mile in 5 minutes, but that is not something I am ever able to do), but also there are life contingencies that limit or make less accessible the possibilities of choice (viz. the geographical location of the person or her socio-economic context). That is why Raz asks for availability of options that enable the person to choose in the long- and short-term timespans, and for varied options, in a logic of quality of diversity above quantity.

Regarding independency, this is a requirement that matches to a certain extent Dworkin's idea of procedural independence.[17] Dworkin is a proceduralist, and seems to believe autonomy is satisfied if the person is not significantly influenced in the process of choosing. Raz goes deeper, and distinguishes—in order to assure independence—that the person should be free from coercion and manipulation (an idea to which contract law seems to adhere by the doctrines of mistake and duress). Coercion is the removal of some or all the options from the person's horizon of choice, while manipulation is the perversion of the process of choosing. Independence as a condition for autonomy is justified in the need to ensure the authorship of a life is truthfully self-conducted: the person-author must have all the adequate perspectives to formulate her desires and must do so relying for the most part in her own judgement.

E. Integrity

It has been claimed so far that the autonomous person authors her own life, being emotionally and intellectually capable of committing to a sufficient variety of long- and short-term choices, which she will consider without uncalled external interferences nor impositions. However, something might be missing, what I will call the condition of integrity.

Integrity is here used in the sense of stability of the person as author of her own life.[18] When one assumes that autonomy has as a major implication and at its essence the ideal of self-authorship, one can claim that a certain degree of planning or stability of principles is required. The person who refuses to give one second of

[17] Dworkin (n 1) 18–19, 21.

[18] This notion is not the same as that of political integrity used by political philosophers who engaged with R Dworkin's *Law's Empire*. See, in particular, G J Postema, 'Integrity: Justice in Workclothes', (1997) 82 Iowa L. Rev. 821, 825. Integrity is presented there as a feature of systematic coherence of the legal system that can justify the rule of precedent. Contrarily, I align with the account of integrity given by C Korsgaard, *Self-Constitution: Agency, Identity, and Integrity* (Oxford University Press 2009). Korsgaard values integrity as an ideal of unity of being, composed by a unity of will and of coordinated agency directed by the unity of will towards a unity of end that drives all the being.

thought before acting is most probably acting on instinct, rather than authoring her life, because there is normally no attitude of finding and refining the self when acting on instinct. The author, it has been pointed out, is generally aware that choices have impacts and knows that by choosing he is adhering to the outcomes of his choice. In sum, the idea of authoring a life implies that the person constantly legitimises her life events as part of the life she accepts and seeks. Thinking of tapestry weaving as a metaphor, the weaver chooses a strand or accepts one that is given to him, consequently weaving it into his unfinished work until the whole of strands makes sense as a tapestry. Such is the work of the autonomous person that, much like the weaver, does not insert a strand by chance—even if the strand was not picked by him, he will make sense of it applying his artisan craft and personal touch to include it in the tapestry.

Some may criticise integrity as too strong a condition. It can be argued that asking for a person to have integrity is to deny the possibility to change one's mind. However, integrity does not mean absolute unity of life and is reconcilable with a change of mind. If integrity were a requirement for unity of life, it would impose a premature and permanent adhesion to a life plan. A person would need to author her life at once and to stick to the original project. Contrarily, what has been defended is that authorship implies a discovery and refinement of the self through renewed choices, under the conditions for autonomy. Indeed, integrity may even require that the person boldly changes her mind, when some choices reveal that the self-author is distancing himself from his previous outlook on life. Again, life is authored autonomously through short- and long-term options, and the autonomous author has to be free to play with his choices as to give significance to his self. This implies that the unity of one's life may never exist, for the person will mature, accepting some new ideas and disregarding old ones, and even at times rehabilitating old options. Integrity respects the flexibility necessary for the autonomous life and simply asks for awareness of the whole. Thus, requiring integrity is no more than making concrete the idea that autonomy is a holistic virtue, to be verified in each specific choice and in the general trend of one's life.

There may be a tendency to consider integrity as a separate value, even though correlated to autonomy.[19] It could as well be a competing value to autonomy, in need of some harmonisation of values and eventually accepted as a higher or lesser value when contrasted with autonomy. I think that is overcomplicating the issue. If integrity is umbilically associated with autonomy as I defended, the conciliation of both ideas is facilitated; if it is an independent value, one will always need to prevail over the other, or both will be equally affected. In the case that autonomy is placed at a higher level, it would follow that an autonomous person could very well be unfaithful to herself (something like an anti-author, an

[19] Dworkin (n 1) 32.

individual set to dismantle any authorship choices), and autonomy would be no more than an ideal of legitimised incoherence. In the reverse case of integrity being placed at a higher level, it could follow that the autonomous person would never be allowed to change her outlook on life or, at least, that short-term options would be discouraged in favour of long-term commitments. Were the two at the same level, yet independent, we would enter into a paradoxical debate between the ability to change one's mind purported by autonomy and the necessity to be faithful to oneself purported by integrity. It seems easier to reconcile both according to the understanding that integrity in a weak sense is a condition to an autonomous life, as proposed above.

IV. Responsibility and Autonomy

From the concept of autonomy here defined, one can draw a consequence of the autonomous life, that of holding the autonomous person responsible. Quante treats autonomy and responsibility as concepts of reflection, that cannot be explained in an isolated form and demand their interplay to be taken into consideration when explaining either one.[20] My claim is that the autonomous person can be ascribed responsibility for her autonomous life, i.e., can be confronted with the choices she makes and the life she chooses to author. This is arguable on the assumption that a person can only be held responsible for that over which she has power; or, in different terms when applied to choices, a person can only be held responsible for choices she made autonomously. Autonomy emerges then as a power; the individual, when acting with autonomy, has power over his choices. With power comes responsibility, for those who have control will have the consequences of their actions imputed to them. In this sense Quante affirms that 'a subject is the adequate addressee of an ascription of responsibility only if she has command of the properties and capacities to make her a rational decision maker', which for him means acting with autonomy. [21]

Contrarily to integrity, we are not stuck with irreconcilable values, because autonomy and responsibility coexist without being paradoxical. The autonomous person can always be held accountable for her autonomous choices, and the responsible person can certainly be autonomous in her choices. Responsibility should not detract autonomy, out of an idea of 'fear of freedom',[22] because a person cannot be truly autonomous if she is not willing to accept the outcome of her choices. In this sense, someone who avoids choices because choices come

[20] Michael Quante, 'Being Identical by Being (Treated as) Responsible' in Michael Kühler, Nadja Jelinek (eds), *Autonomy and the Self* (Springer 2013), 254.

[21] ibid 260–261.

[22] Dworkin (n 1) 67.

with consequences is autonomous in a very poor sense, considering that avoiding choices is part of that person's autonomy. After all, if in being autonomous she is authoring her life, she has to be satisfied with the life she is shaping with all its consequences.

In fact, responsibility works in favour of autonomy when seen as a deliberation-aid, a factor taken into consideration when a person is about to act as to author her life. On the verge of choosing, the person considers all the adequate alternative options and assesses their implications. Implications are first assessed against the whole of the life, as to assure their conformity with an attitude of authoring: the person questions whether by choosing X over Y on the long-term she will keep in line with previous trends, inaugurate new perspectives on life or even reformulate old tendencies, or whether by choosing W over Z in the short-term she will be coherent with her system of beliefs or her world view that inspire the personality and life she is authoring. This is mainly the concern of integrity. Responsibility, on its turn, adds to these considerations by making the person aware that there are inner and outer implications to her actions, on the way her life is being shaped and on the way others perceive and relate to her life. It makes the autonomous action or decision not only a matter of the inner realm, but relevant in terms of accountability, which implies that there is true autonomy when the person accepts being held responsible.

Notwithstanding, responsibility seems to be opposed to the possibility of changing one's mind. It is easily argued that holding someone accountable will impede the abrogation of a past decision or the adopting of a new attitude towards life issues. In our case, forcing A to keep the lease prevents him from buying a house. This perspective is false. If anything, responsibility is a strong device to enhance the chances to autonomously change one's mind. By holding the person accountable, one is ensuring not only critical self-reflection at the moment the person decides or acts, but also at every subsequent moment. A's liability for breach of contract is what translated this view on responsibility to our case. Thus, responsibility addresses concerns about arbitrary changes of mind, by preventing them: the autonomous person will ponder consequences when considering new perspectives for her life, and by doing so the procedure of autonomy is aggravated. Moreover, responsibility offers mechanisms to change one's mind by identifying the consequences that may oppose or prevent such a change and allowing the autonomous person to deal with them. Let us imagine a child faced with the option between studying and playing. At a first moment, the child opts for study. Later on, the child starts to consider the willingness to play, and ponders whether there would be any consequences to stop studying. We can admit the child has intellectual and emotional maturity to know study and play must be balanced. If the child is aware that playing at that

time will directly impact a grade on the next day's test, and still decides to play, the child is being both responsible and autonomous. There was acceptance of the consequences, which were seriously considered, and an identification of the worth of playing with the life the child is authoring. The example is naïve, naturally.

V. About the Paradox of Autonomously Promising and Autonomously Changing One's Mind

The need for the thick concept of personal autonomy I propose above, one which may be of use in explaining the phenomenon of contract as promise, is strictly linked with the paradox of promising and changing one's mind for the autonomous person. In this Chapter, I will start by debating the worth of both promising and changing one's mind in face of the ideal of personal autonomy. Analysing the paradox will put to test the thick concept of autonomy explained above, assessing whether it offers a way out of the paradox while remaining consistent with the ideal of personal autonomy.

A. Understanding promises in general

Kimel defines promising as 'the practice by which people voluntarily undertake obligations to others'.[23] Along the same lines, Hogg says that 'a promise is a statement by which one person commits to some future beneficial performance, or the beneficial withholding of a performance, in favour of another person'.[24] The necessary elements of a promise, verifiable in any promissory event, are the assumption of an undertaking (commitment and obligation may or may not be interchangeable concepts) by a person to the benefit of another. By promising, the promisor is agreeing to act (or omit an action) to the benefit of the promisee, which can or not be the sole beneficiary of the promise. Regarding the benefice, it is possible to say that the promisee is always a beneficiary, even if the only benefit he acquires is the fulfilment of the promise itself to the immediate benefit of a third person or the promisor himself (v.g., a mother can ask her son to promise her he will quit smoking, and in fulfilling that promise it is clear how both benefit: the son, in his health, the mother, in her motherly concern for the well-being of her son). Recalling our case, A promises to lease from B, undertaking to pay rent for the period of three years, and in this case B is beneficiary of that promise, a benefit that translates into receiving rent. Reversely, B promises to lease to A, undertaking

[23] Dori Kimel, From Promise to Contract: Towards a Liberal Theory of Contract (Hart Publishing 2003) 27.

[24] Martin Hogg, *Promises and Contract Law: Comparative Perspectives* (Cambridge University Press 2011) 6.

to allow A to live at the property, and A is beneficiary in that B cannot pursue an unmotivated eviction.

The voluntariness of the promise is, in my view, an incidental element of the promissory events, although a commonly perceived feature. Ideally, the assumption of obligations should be done voluntarily, but life provides us with various situations in which obligations end up being imposed without much questioning as to their validity (v.g., the parental obligations that come from parenthood, to which no one expects the parents to assent). The language of the promise involving an undertaking or obligation is justified on the accountability that normally comes with a promissory event: what distinguishes a promise from a resolution is, indeed, the fact that someone will be able to demand the fulfilment of the promise. The promisor enters a commitment, whilst the person who makes a resolution simply expresses the desire for the adoption of a conduct that, in her failing to do so, will not defraud an expectation. No doubt the promisee can also be coerced to accept the undertaking, or may be surrounded by circumstances that impede his freedom. The fact that the promise is entered into without voluntary assent by the promisor does not, in my view, change much of its character: rather, it only shows that there are voluntary as involuntary promises. However, the value of these is different: it will be readily accepted that justice demands caution when it comes to the enforcement of involuntary promises to the point of accepting that the promisor does not conform to his commitment. This is why I did not include enforcement as an essential element of a promise, because enforcement and accountability are in different levels: the former implies mechanisms that tend to provoke the fulfilment of the promise, whereas the latter merely implies that someone (normally, the beneficiary or promisee) will be expecting fulfilment of the promise and seek to enforce it, even though it may be unfair.

B. Promises and personal autonomy

Promising is, foremost, a social practice.[25] There may be various reasons for promising, ranging from external to internal, from egoistic to altruistic. The qualification of a promise as a contract adds, to whatever these reasons are, the wish to enter into a contract, or better, to create a legal relationship that covers the promissory relationship. It is not only that B wants to make some money by leasing the property, B wants to receive the money within a contract that sets the rules in place. Notwithstanding, these are all motives, and personal autonomy is

[25] Many attempts at understanding promising rely on such an assumption. For instance, Fried (according to the account by Kimel (n 21) 10) centres his explanation of the nature of promising in the fact that it is a social practice, and Hogg's definition of promise ((n 22) 4) starts from the idea that an empirical perspective can help best elucidate what is a promise.

both motive and justification. People promise because they can, and their ability to do so is—outside of coercion or external impositions—a virtue of their autonomy as individuals; also, the value of the ability to promise is normally grounded in autonomy.

Autonomy as a motive for promising seems to compete with other motives: one cannot exclude that human beings may be tempted to commit themselves merely on the grounds of their ability or need to do so. This is a line of enquiry of difficult pursuit in a paper like this one, which does not cover behavioural sciences.

More relevant to my enquiry is autonomy as an explanation and justification for promising and its worth. I do not intend to explore other justifications of promises and their worth, for there may be many moral aspects that justify it (e.g., the claim that trust is valuable and can be created by a promise). Promising is valuable as an autonomy-maximising device,[26] [27] assuming that what maximises a good is in itself good. The ability to commit oneself through a promise is almost an inevitability of authoring one's life autonomously, for without the ability to promise the range of autonomous choices would not be complete. As argued above, personal autonomy is an ideal of self-authoring that requires the possibility of committing to a sufficient variety of long- and short-term choices. Promising, as claimed now, is a mechanism of self-commitment that presents the possibility of making short- and long-term undertakings binding to the promisor. The connection between both in this aspect is almost immediate: promising is one of the modalities of authoring one's life in an autonomous way, broadening the adequate range of options in quality and quantity, naturally only when promising respects the others conditions for autonomy. Moreover, promising gives origin to solid commitments, which is in line with my claim that integrity is a condition for autonomy. Binding oneself through a promise is an exercise of integrity that indicates that the promisor acknowledges the desirability of a reinforced form of commitment, for which he will be accountable. Indeed, the accountability of the promisor shows the soundness of his self-authoring endeavours, because by

[26] Fried (according to Kimel (n 4) 1) does not use the expression and talks in terms of freedom: 'In order that I be as free as possible … it is necessary that there be a way in which I may commit myself'. On these grounds, Kimel says Fried's idea is that of promising as a freedom-maximised device, from which I borrowed the concept to the expression used. Freedom and autonomy have blurred conceptual borders, which would be hard to trace here; therefore, I have avoided the use of the word freedom.

[27] Accepting promising is an autonomy maximising device does not equate to accepting that indiscriminately promising will ensure or favour autonomy. Quite the contrary, as Kimel notes ((n 4) 3).

accepting a commitment the promisor is authoring his life with some degree of permanence.

C. Autonomy and the possibility of a change of mind

A change of mind can take a multitude of forms: it can mean adopting a totally new idea, rejecting an old conception, updating a previous understanding of reality, conforming a taste to novelty *etc*. What seems necessary is that the person transitions from a mental state of affairs to another, rearranging the factors that will be determinant in making self-authoring choices. Thus, a change of mind in this case is not solely concerned with the world of ideas or representations; it looks at the person's mentality shifts that are translatable into new opportunities for self-authorship.

The concept may come up as hermetic, but what concerns us is the autonomous person and the ability she has of readjusting the train of thought that lead her to previous choices or will condition future choices. Truly, any intellectual or emotional phenomenon with the potential of impacting choices can be called a change of mind for this effect. One of the most impressive illustrations of a change of mind is the acquisition or development of maturity while growing up: almost every person, during the transition from childhood to adolescence, and then from adolescence to early adulthood, will contact with new ideas that replace old ideas and with new realities that will require the person to form an original (as regards her mind) opinion.

The change of mind is a virtue of the autonomous person that, by itself, can occur outside the conditions for autonomy. It cannot be required that the transformation of the mental state of affairs of a person occurs only with respect for the conditions for autonomy, for the autonomous person can have no control over many of the life events that imprint new perspectives in her mind. In fact, it could be argued that most intellectual and emotional events that provoke a change of mind are exogenous, attributable to other persons or to general life circumstances rather than to spontaneous and critical self-reflection ordered at a change of perspective about life. Additionally, changing one's mind is not required by autonomy, considering that in theory (and maybe in practice, although examples all seem to be open to exceptions) a person can have a deeply entrenched system of beliefs and a rooted perspective on life that will not be affected by any kind of external event and that, when subject to critical self-reflection, will invariably end up unaltered. It is thus relevant to add the words "possibility to" to the expression

"change one's mind", in order to emphasise that this is something that can potentially but does not need to occur.

Within the framework of an autonomous person's life, assuring the possibility to a change of mind is essential. It is required by the condition that autonomous choices can only arise from an adequate range of options, which will positively be broadened by facilitating changes of mind. The adequate range of choice is primarily verified (as to its fulfilment as a condition for autonomy) according to the state of affairs presented at the moment the person will choose. But the possibility of a change of mind brings to the moment of choice potential future alternatives that would be impossible to consider if it were not for a change of mind being allowed. That is, when confronted with the moment of choice, the person will be able to decide with autonomy not only because there is sufficient quality and quantity of alternatives, but also because the future alternatives that are somehow incompatible with present choices are available through the "scape valve" of a change of mind. Furthermore, accepting changes of mind removes, from the perspective of the author of his own life, an autonomy-impairing fact: the 'fear of freedom'.[28] This 'fear of freedom' is, in fact, fear of the implications of acting with autonomy. It is expected that a conscious person acknowledges the future implications of her choices, and sometimes that can be an obstacle to autonomy, when the person prefers not to choose or decide in order to avoid making the wrong choice. Allowing a change of mind to be in the perspective about the future of the autonomous person reassures her that, in general, a bad decision is potentially revocable or changeable, which prevents the crisis over having to choose in an all-or-nothing attitude.

D. Unveiling the paradox

The relationship between autonomy and promises is apparently paradoxical. On the one side, there is the view that the ability to promise maximises autonomy. On the other, promising appears as a commitment that impairs autonomy by holding the promisor accountable and thus preventing that at a later stage he prefers an alternative choice to the one he made when promising. It seems as if promising has to be amongst the person's options to enable an autonomous choice and that by fully committing herself she is benefiting from an autonomy-maximising device, while in fact she is forfeiting the possibility to change her mind: the paradox tells us that the autonomous person would be paralysed by the fear of freedom.

There would be no paradox if promises were easy to disregard. That being the case, promises would seem to be fully in accordance with the thin ideal of

[28] Dworkin (n 1) 67.

autonomy and there would be no place for a 'fear of freedom'. However, we see how promises are backed by enforcement mechanisms that at least highly discourage non-performance of the promissory undertaking. Are these totally unjustifiable in an account of promising relevant for the autonomous person? The answer is to be found in the condition of integrity and in the value of responsibility. Both ideas give good reason to accept promising as an autonomy-maximising device. Integrity, asking that the autonomous person exerts autonomous choices contemplating the whole of her life and the necessity to be truthful to and consonant with the self she is seeking to author, allows us to view promises as one of the preferred ways of shaping one's life. Autonomy that respects the condition of integrity is present in many modes of choosing, but the choices included in a promise respect it more patently. The accountability that the promisor accepts reveals his level of reflection in the whole of the life he is authoring, by being sufficiently sound in principal that a more permanent level of commitment does not scare him away from living the autonomous life. As regards responsibility, it is worth recalling that it enters with autonomy in a list of relevant values, thus allowing a claim that if, for one side, autonomy may ask for a greater level of flexibility on one's commitments (especially considering the possibility of a change of mind), for the other it is not an absolute value and needs to be harmonised with other equal worth values. On the harmonisation between these, I concluded that responsibility works as a deliberation-aid to the autonomous person, a moral side-constraint, considering particularly that true autonomy exists when the autonomous person willingly accepts responsibility. Consequently, given that promising is a way of willingly accepting responsibility for an undertaking, it is easily inferred that promising maximises autonomy and the fact that a person cannot get out of a promise without a consequence is actually an encouragement to choices having to be made with a high degree of autonomy. I am lead to conclude that the paradox of promising being at the same time in favour of autonomy and a limit to it is apparent, for autonomy itself agrees with the limitations to one's life imposed by the entering into promises in virtue of the condition of integrity and of its interplay with the value of responsibility.

VI. Autonomy and Contract Law

So far, I have explained how I see personal autonomy, its conditions and interplays with responsibility, with hints at contract law. Additionally, I have sought to open the debate about the nature of promises and to give an autonomy-informed account of the act of promising, dealing with the most pertinent problems that promising poses to the autonomous life. These were not "diversion manoeuvres"

to avoid the line of enquiry I chose to follow in this paper: they already addressed many of the concerns about contract as promise in the context of an autonomous life. Simply, the debate required a certain degree of background assumptions that had to be formulated and tested. In this chapter I will finally offer a reading of contracts informed by the ideal of personal autonomy.

E. Contracts and the maximisation of personal autonomy

If anyone was asked what institution of law conforms the most with the ideal of personal autonomy, advancing contract as an option would be a safe bet. Contracts are, by essence, a legal institution destined to fulfil the individual aspirations of conforming one's life within a legal binding framework.

Within the law of contracts, freedom of contract has a special place as an autonomy-inspired aspect. If the capacity to self-impose an obligation is a 'logically-antecedent freedom',[29] the idea of freedom of contract is a first and most expressive reflexion of that capacity in the legal realm. Freedom of contract is a principle or idea accepted by most jurisdictions that implies the liberty in deciding whether or not to enter a contract, the independence to choose what type of contract to celebrate and the possibility to conform the contract through the inclusion of some provisions and the exclusion of others. Its importance for a legal system is evident in the Draft Common Frame of Reference,[30] which phrased it as a main principle:

> II.—1:102(1). "Parties are free to make a contract or other juridical act and to determine its contents, subject to any applicable mandatory rules."

Along the same lines, the Art. 0:101 of the *Principes directeurs* places freedom of contract as a major governing principle of this prospective European Private Law:

> Each party is free to contract and to choose who will be the other party. The parties are free to determine the content of the contract and the rules of form which apply to it. Freedom of contract operates subject to compliance with mandatory rules.

There can be multiple reasons for the relevance of freedom of contract, such as economic and social factors, but its role as an autonomy-maximising mechanism

[29] Kimel (n 4) 1.

[30] Von Bar Commission, 'Draft Common Frame of Reference' <https://www.law.kuleuven.be/personal/mstorme/DCFR.html> accessed 03 March 2019.

is what matters for the effect of this paper. By opening up the contractual practice in its formation stages to various alternatives (negotiation and effectively entering into a contract), freedom of contract assures the fulfilment of the conditions of autonomy from the moment the parties are called to make their first contractual decision (that of starting negotiations that may lead to the formation of a contract). Firstly, there is adequacy of options both in quantity and quality: the parties can decide what contract to celebrate, adapt some of the legal terms and reject others, introduce mutually agreed provisions; previous to that, they can decide whether or not to start a negotiation, and in the end—even after the whole of the negotiation had been done—they can withdraw before accepting the final terms (subject to possible consequences, that are of very exceptional nature, such as situations of trust leading to pre-contractual liability). In our case, it is open to A and B to enter into a lease or not, to decide the moment they start negotiating, the type of contract they will celebrate (B could sell instead of lease) and, ultimately, until there is effectively a contract both A and B can say they are no longer interested in entering into a lease. Secondly, the kind of deliberative procedure that freedom of contract causes favours the condition of integrity: by opening up so many potential choices, there is an increase in the parties' awareness that there are many possible solutions to their deliberation, deeming the choice ultimately made as a more matured and well-thought through one. Recalling the lease, A has the time to consider the commitment a three-year lease represents in terms of paying rent and may decide to employ the rent money to obtain a loan and buy some property.

Additionally, even though not immediately related with freedom of contract, contract law includes provisions that assure sufficient maturity (viz. minimum age to independently exercise some rights) and the absence of factors vitiating the will of the parties (namely by declaring void contracts entered into without independent deliberation provoke by external interferences), thus ensuring the emotional and intellectual capacity of the parties. In this matter, the doctrines of mistake and duress are of particular importance and evidence how the law is concerned with having an autonomous decision to bestow upon it the binding force of contract.

It may be questionable whether contract law has as its main reason the protection of the parties' autonomy. Some might argue that freedom of contract militates in favour of such conclusion, namely on the grounds that all contract law can be read as either a result or an exception to freedom of contract, which in turn is an autonomy-maximising device. That is, the rules that preside over contracts are seen as forced by the demands of freedom of contract or as legal restrictions to the scope of said freedom, in a way that all contract law revolves around freedom of contract. We may call this perspective the maximalist view of freedom of contract, and immediately point out that it neglects other ends and functions of contract law such as the facilitation of economic transactions and the

prosecution of other types of policy. In my view, these other ends and functions serve to show that freedom of contract is not the main value within contract law, but certainly one of the most relevant amongst others (especially the facilitation of economic transactions between individuals). That being the case, I would argue that personal autonomy is amongst the most valued ends prosecuted by contract law, in equality with some others already mentioned.

Even though personal autonomy is not the main reason for contract law, the fact that it is one of the most relevant ones implies that a reading of contracts within the framework of personal autonomy is reasonable. As I started by delimiting in this paper, autonomy has to be able to justify the ability to be bound by contract, while solving the paradox that an autonomous person bound by contract will not be able to easily change her mind. Again, the other values prosecuted by contract law can have answers to these questions: v.g., the need for protection of economic transactions can justify the ability to be bound by contract and the limitations to the possibilities of leaving a contract. However, in spite of all other alternative values, autonomy must answer the concerns raised by contract law. Otherwise, the value of contract would be undermined, at least to those who most directly benefit from this institution (the parties).

F. Challenges that contracts pose to the ideal of autonomy

In Chapter II I explored the assimilation of contract as promise proposed by the doctrine of the same name. The main conclusion was that contracts are qualified promises, i.e., promises to which the law confers relevance by attaching to them legal enforcement mechanisms. Kimel seems to refuse such *tout court* assimilation, at least to a certain extent, when he affirms that seeing contract law as a system of enforcement of promises can lead to a situation in which promises are enforced outside the prevalent 'nonlegal habitat for promissory activity'.[31] I do not wish to contend the idea that promises are surrounded by a particular environment that offers norms (namely, for enforcement) that can be different than those offered by contract law. Notwithstanding, the problems identifiable in promises are in many ways similar to those posed by contracts when reality is read under the light of personal autonomy.

A first aspect is that contracts, or rather, the possibility of entering into a contract, are a demand of personal autonomy. The ideal of personal autonomy as the self-authoring of one's life requires that there is at least one legal institution presenting the individual with a means to self-regulate aspects of his life, especially when rightly understood that self-imposition of obligations can be an autonomy-

[31] Kimel (n 4) 18.

enhancing feature. Contracts, much like promises, create an obligation to an undertaking that is grounded on the agreement of the parties, promisor and promisee. The existence of the possibility to enter into a contract significantly increases the quantity and quality of alternatives of choice, which then reflects as the improvement of one of the conditions for autonomy (much of what I said regarding freedom of contract applies here). Claiming that contracts are a demand of the ideal of personal autonomy may be a stretch, for other alternative commitment methods could exist (e.g., promises in general, without legal framework, could perform the same autonomy-enhancing function as contracts). Nonetheless, the importance of contracting for the autonomous life is less connected with the absence of alternatives than with the necessity that the autonomous person finds a specific alternative in the law, to which she has a legitimate expectation considering that entering into legally relevant relationships is a way to author one's life valuable to many individuals. In our case, A, looking for a house, can either buy one, rent one, or borrow one from a friend. The law of contract offers three contractual devices (in fact, many more) that allow A to achieve his goal. A would not be autonomous, he would not be able to live an autonomous life deciding where to live and how to get a house, if contract law did not offer the contractual types of buying, renting and lending.

A second aspect is the binding force of contracts, which seems in disagreement with the ideal of personal autonomy. Under normal circumstances, a contract is celebrated in view of the fulfilling of the contractual promise, and for that effect it is endowed with legal enforcement mechanisms. On one side, this is a feature that enhances autonomy, because it provides a binding alternative of choice for those who seek it. The autonomous decision of A to rent an apartment is truly autonomous because he can get a lease for three years, allowing A to plan and author his life, which would not be as easy if no guarantee of having a house was in A's horizon. On the other side, entering into a contract poses the same problem as promises: it seems that the promisor (party to the contract) is bound by his promise without an alternative or the possibility to change his mind. In fact, a change of mind concerning the suppositions the person took into consideration to accept the bargain is normally deemed as irrelevant for the effect of contractual validity (which is assessed at the moment of formation of the contract) or getting out of the contract. Most of the legal mechanisms that allow one party to get out of the contract are rooted in the value of autonomy: e.g., the doctrine of frustration opens an exit in the contract on the grounds of a change of events, and not a change of mind. As I defended regarding promising, accepting the condition of integrity helps us realise that the level of compromise which arises from a contract is consonant with the value of personal autonomy. Contracting is, perhaps to a

higher degree than promising, a privileged way of showing soundness of principle when acting autonomously, for by entering into a contract a person is deliberately assuming a legally protected commitment that is thereafter part of her obligations both in a moral and in a legal sense. Moreover, contracting is also a privileged way of assuming responsibility in an aggravated way, because a breach of contract brings about serious consequences for the party in breach.[32] In our case, B decided to rent her house; she could have sell it, but she preferred the commitment to a tenant for three years in exchange of rent, and she was willing for such an engagement. When the lease was entered into, B had the three years in mind and knew for that time the bargain would be mandatory. Her soundness of principle, showed in her willingness to be a landlord, is what allows for a three-year lease.

The fact that entering into a contract apparently creates a situation that does not enhance autonomy can be contested not only on the grounds of the condition of integrity or the value of responsibility for the autonomous life, but also on features of contract law that mitigate the apparent harshness of the binding force of contracts. Having the Common Law in mind, one must be aware that specific performance is not the mainstream remedy for a breach of contract, which serves to show that there is not as much pressure as probably thought in obtaining the fulfilment of the contractual promise. It is possible to see damages (especially clauses foreseeing liquidated damages) as a price to leave the contract, which is understandable as opening up the option to the autonomous person to exit the bargain in a fair manner, respecting at the same time the binding force of the contract and her change of mind regarding her bond. The law offers alternatives to pure specific performance of the contractual promise, and these can be flexible enough to accommodate the promisors need to exit the contract caused by a change in his mind. If B, in our case, would like to sell the house after one year, she can; it is just that contract law makes it harder, not impossible, to change one's mind. Naturally, these "exits" may have a cost (apart from the money-value of damages, one can envision the loss of the counter-promise or of the promisor's credibility in the market as *inter alia* costs of fully embracing a change of mind), but accepting the costs of one's decisions is a virtue of the autonomous life.

Again, integrity and responsibility explain a lot when talking about contracts and a change of mind. The law is keen in allowing contracts to exist and flourish, and the law has contractual performance in view. This is, I believe, because contract law considers the legal person entering into a contract someone

[32] I hope in these last two paragraphs to have addressed Gutmann's idea that there is some sort of trade-off 'between autonomy, conceived as an individual right, on the one hand, and the ideal of autonomy, as a consequentialist, summable individual value and public good, on the other hand'; cf Thomas Gutmann, 'Theories of contract and the concept of autonomy' (*Preprints and Working Papers of the Centre for Advanced Study in Bioethics*, Mü nster 2013/55) 21.

gifted with integrity and responsibility. Contract law offers a framework for the person to develop her life plan and to author her life, and this framework includes long- and short-term options that bring unity to the legal person within the legal realm. That is why A, after renting, can contract a housecleaning service: the diversity of his contracts show the flow of autonomous decisions A has made has an author of his life. In addition, responsibility for the autonomous person means legal accountability for the legal person. Entering into a contract always implies (or should, ideally) having in view the consequences of that contract, both while it is performed as after termination. The truly autonomous person is not entering into a contract light-heartedly: she considers the costs of being in a contract and of leaving that contractual relationship. She is responsible and therefore does not fear accountability.

VII. Conclusion

I hope that my reading of contract as promise has served my two purposes: as an illustration of the interplays between the ideal of personal autonomy and private law by reference to one of its most relevant institutions (contracts), and as a useful account of how promising (and therefore contracts) conform to that ideal.

Personal autonomy is an ideal of self-authorship that assures that, by autonomous choices, the person is shaping and refining herself. Autonomy is subject to certain conditions: the intellectual and emotional fitness of the person; the existence of alternatives varied in quantity and quality; the absence of external constraints on the person; and a minimum soundness of principle in the process of choosing. In any case, autonomy is not an absolute value, and it must be balanced with other values of equal worth, like the value of responsibility, which asks for accountability of the autonomous person for her own autonomous choices.

Promising, as an assumption of an undertaking, can enhance autonomy while at the same time jeopardizing it by preventing (in some circumstances) a change of mind. This is not paradoxical, considering that autonomy requires integrity and that responsibility mitigates the power of the possibility of changing one's mind. The same logic applies to contracts, especially because they are considered as qualified promises (i.e., promises to which the law confers legal mechanisms of enforcement). Maybe due to the enforcement of contracts appearing as more grievous, entering into a contract can seem a total forfeit of autonomy. This is not absolutely true, considering that contract law offers ways out of a contract, which have their costs.

The autonomous person has, at last, a way out of the paradox that paralyses her out of fear of freedom: she can conform to a life plan that takes into account

her past and future choices, being as more autonomous as she is more in line with her own aspirations and the life she is authoring; at the same time, the autonomous person knows that with autonomy comes responsibility, meaning that any decision has consequences and choosing something is also choosing its consequences and to be accountable for them. In our case, A and B are truly autonomous when they know a contract will, except for situations of frustration, duress or mistake, bind them for three year, with obligations to each party, and that to get out of the contract before the time the consequences of rescission will apply.

The Scope and Legal Effect of Choice of Law in International Arbitration

Dominic Npoanlari Dagbanja[*]

I. Introduction

The scope and legal force of choice of law have not been given relevant attention in both legal scholarship and international arbitration. This article is concerned with such issues. I seek to establish that choice of law limits the applicable law to law that is freely, voluntarily and legitimately chosen by the parties, and. arbitral tribunals do not have unfettered discretion to administer what they may term justice outside the scope of that law. Cross-border commerce and uncertainty as to its applicable law are features of today's world economy.[1] Thus, contractual parties may, as they commonly do, choose a particular law to govern their legal relationship. They make a choice of law in the exercise of their autonomy and freedom of choice to do so. Particularly in the context of international business transactions involving parties from different jurisdictions, if the parties do not make a choice of law, uncertainty may arise as to the applicable law. This can

[*] Lecturer in Law, The University of Western Australia Law School, Perth, Australia; BA (Hons), University of Ghana; LLB (Hons), University of Ghana; Qualifying Certificate and Certificate of Enrolment, Ghana School of Law; LLM, University of the Pacific, USA; LLM, The George Washington University, USA; PhD, The University of Auckland, New Zealand; Graduate Certificate in Tertiary Teaching, The University of Western Australia, Australia. Parts of this article, especially the case analysed, came from my PhD thesis under the supervision of Professor Jane Kelsey and Associate Professor Christopher Noonan. I would like to thank them for their feedback and support. I would also like to thank Professor Dan Priel of York University Osgoode Law School and Professor Paul Beaumont of the University of Aberdeen School of Law whose comments helped me improved the quality and clarity of this article. Email for correspondence: dominic.dagbanja@uwa.edu.au.

[1] Sagi Peari, *The Foundations of Choice of Law: Choice and Equity* (Oxford University Press 2018) xv–xvi.

make it difficult for the parties to know and comply with the appropriate law in the course of contractual performance or when a dispute arises.[2]

The uncertainty as to the applicable law in the absence of choice of law may arise due to at least two reasons. Firstly, the applicable law to a contract is determined by applying rules of private international law belonging to a particular national legal system. Different countries have different legal rules of private international law for the determination of the law applicable to a contract. Thus, unless a choice of law is made by the parties, it may be difficult to determine the law applicable to the contract on the basis of rules of private international law within national jurisdictions. Secondly, uncertainty may still arise even if the application of rules of private international law to determine the applicable law are certain. This is because those rules may be too general or vague to enable a reasonably certain and accurate determination to be made. Choice of law then, when made within the legal limits of rules of the relevant system of private international law, can lead to certainty and predictability as to the law applicable to the contract.[3]

As stated by the United Nations Commission on International Trade Law (UNCITRAL), the autonomy of the parties may be limited to the extent that they are permitted to choose a legal system only if it has some connection with the contract.[4] This may require a choice of the legal system of the country of one of the parties or of the place of performance or even of the seat of arbitration.[5] It is also possible for the parties to choose the law applicable to the contract without these restrictions.[6] To avoid these uncertainties surrounding lack of specificity as to the law applicable to the parties' transaction, the parties choose the law of a particular country to govern their contract.[7] Their "very purpose in specifically selecting a law to govern the contract is to settle in advance all doubts as to what internal law is to be applied."[8] Making a choice of applicable law also serves the interests of the international community or society at large because, as asserted by Derrick Wyattc, the choice of a particular law can further a policy favoured by

2 Elliott E Cheatham and Willis LM Reese, 'Choice of the Applicable Law' (1952) 52(8) Columbia Law Review 959.

3 UNCITRAL, *Legal Guide on Drawing up International Contracts for the Construction of Industrial Works* (New York: United Nations, 1988) 299–301.

4 ibid 301.

5 ibid.

6 ibid.

7 ibid 300.

8 Columbia Law Review Association, Inc, 'Conflict of Laws. Choice of Law in Contracts. Intent of the Parties. Renvoi.' (1940) 40(3) Columbia Law Review 518, 523.

most commercial nations since such nations allow citizens a degree of freedom of contract.[9]

Based on a choice of law analysis, I seek to contribute to an understanding and appreciation of the continued role of municipal law (the law invariably chosen by the parties) in the regulation and protection of foreign investment, trade and other international business transactions. I will do so by examining the limitations of the choice of law on the scope of applicable law, the rights of the parties to international business transactions and the powers of arbitral tribunals. According to UNCITRAL, if a dispute "is settled in arbitral proceedings, the law chosen by the parties will *normally* be applied by the arbitrators."[10] This attests to the fact that arbitral tribunals do depart from the choice of applicable law made by the parties, and in fact as shown in Part III, arbitrators have rejected the parties' choice of municipal law. This seems to be at odds with the concept of arbitration, which is founded on the consent of the parties. Acknowledging party autonomy and consent as the foundations of arbitration and the central role of *lex loci contractus* in the control of arbitration, the New Zealand Law Commission stated the following:

> Arbitration law concerns a critical balance: the balance which is to be struck between the autonomy of the parties and the law of the land. On the one side of the balance is the agreement of the parties. The parties to a contract or to a dispute agree that their disputes are to be resolved by a tribunal which they establish themselves or to which they agree. The tribunal is to follow a procedure on which the parties may agree, and is to apply the law which they may state. The parties also, in general, pay for the arbitration. That is to say, the whole process rests on the parties' consent and is their creation. But not quite. For on the other side of the balance is the significant weight of the general law of the land. The very agreement that sets up the tribunal is an agreement under some system of law. It is national law, with national courts, which can be used to require a reluctant party to submit to arbitration, and to enforce any resulting award. The law may state the procedure to be followed. The law might, as well, also be used to control the arbitrator.[11]

[9] Derrick Wyatt, 'Choice of Law in Contract Matters—A Question of Policy' (1974) 37(4) The Modern Law Review 399, 408.

[10] UNCITRAL (n 3).

[11] New Zealand Law Commission, *Arbitration* (Report No 20, 1991) [3].

A conflict of law situation arises when two or more systems of law have some possible basis to govern the resolution of a legal dispute.[12] Where the parties have legitimately made a choice of applicable law for the resolution of their dispute, it does seem that no conflict of law situation remains to be addressed because *prima facie* an arbitral tribunal knows the governing law. For example, under the Convention for the Settlement of Investment Disputes between States and Nationals of Other States (ICSID Convention), a tribunal "shall decide a dispute in accordance with such rules of law as may be agreed by the parties,"[13] and may "decide a dispute *ex aequo et bono* if the parties so agree."[14]

The analysis on the legal effect of choice of law is relevant for investment treaty law and arbitration as well, especially because investment contracts commonly require investment arbitration. The backlash against investment treaty law and arbitration[15] is a manifestation of its inability to work to the satisfaction of states. In particular, the limitations and inadequacies of the investment treaty regime call for the need to rethink the role of municipal law, which is invariably chosen as the governing law, in foreign investment and international commercial regulation. The role of municipal law in the regulation of foreign investment and international business transactions depends on tribunals' recognition and acceptance of a choice of domestic law made by the parties in respect of that system of law. I will discuss the circumstances in which arbitral tribunals have rejected a choice of particular domestic law in order to contribute to a new understanding of the continued role of municipal legal systems, international investment agreements (IIAs), and principles of international commercial law in international business and foreign investment protection in light of a choice of law.

International arbitration is not simply a creature of states; it also depends on their goodwill and support through their laws and courts for the enforcement of arbitral awards. Without states' cooperation and support, the international arbitral system will not work. Furthermore, it will operate at risk of its own collapse if it does not respect and uphold national laws simply because these laws will not work in favour of the investor. As Professor William Park argues, freedom from

[12] Jeremy Kirk, 'Conflicts and Choice of Law within the Australian Constitutional Context' (2003) 31(2) Federal Law Review 248, 249.

[13] Convention for the Settlement of Investment Disputes between States and Nationals of Other States (opened for signature at Washington, on 18 March 1965, entered into force 14 October 1966) art 42(1).

[14] ibid article 42(3).

[15] See generally M Sornarajah, *Resistance and Change in the International Law on Foreign Investment* (Cambridge University Press 2015); Michael Waibel and others (eds), *The Backlash against Investment Arbitration: Perceptions and Reality* (Kluwer Law International 2010); David Schneiderman, *Constitutionalizing Economic Globalization: Investment Rules and Democracy's Promise* (Cambridge University Press 2008); Gus Van Harten, *Investment Treaty Arbitration and Public Law* (Oxford University Press 2007).

the constraints of substantive and procedural law is desirable in international arbitration.[16] He further notes that:

> [arbitration] affects not only winners, but also losers, and often society at large as well. The fashions for non-national justice and arbitral autonomy, if pushed too far, will ultimately backfire to compromise the integrity of international dispute resolution. The chemistry of these trends may inflict on the business community an unjust uncertainty even less appealing than the mandatory norms of local arbitration law.[17]

It is, therefore, imperative that arbitral tribunals should respect and uphold national laws put in place by countries.[18] The obligations to respect and uphold national laws arise particularly where the parties have legitimately chosen a particular national legal system as the governing law to their legal relationship.

I argue that where the parties have made a choice of law within the legal limits permitting such a choice to be made, the chosen jurisdiction has the dominant interest to have its law followed. The laws of the jurisdiction of choice must accordingly be applied to the resolution of the dispute and the agreement to arbitrate unless the internal rules of conflict of laws in that jurisdiction point to some other jurisdiction, or if the parties have reserved powers in an arbitral tribunal to do otherwise. Where the parties agree that their *underlying agreement* contains all their understandings, rights and obligations (entire agreement clause) in relation to the subject matter of the contract, the powers of the arbitral tribunal are limited to that agreement, the governing law and the rules of arbitration specified in that agreement. I argue also that where an underlying contract voluntarily and legitimately agreed by the parties expressly indicates that *any* dispute *arising out of* or *in relation to the underlying contract* shall be settled by arbitration and that the arbitration shall be governed by and conducted in accordance with specified arbitration rules, then a dispute as to the validity of the arbitration clause has to be resolved within the terms of those arbitration rules and governing law by the parties in that underlying contract. Furthermore, I would like to point out the emphasis that applicable arbitration rules and conventions on applicable law place on the need for arbitral tribunals to respect and uphold the choice of law made by the parties and the fact that some of the arbitration rules even require that arbitral

[16] William W Park, 'National Law and Commercial Justice: Safeguarding Procedural Integrity in International Arbitration' (1988–1989) 63(3) Tulane Law Review 647, 649.

[17] ibid 649–650.

[18] Daniel Hochstrasser, 'Choice of Law and "Foreign" Mandatory Rules in International Arbitration' (1994) 11(1) Journal of International Arbitration 57, 85.

tribunals shall only make decisions as *amiable compositeur* or *ex aequo et bono* only if so expressly authorised by the parties. In this regard, an arbitral tribunal cannot embark on a journey of its own to do justice according to its sense of equity unless so authorised by the parties. As far back as 1875, Sir George Jessel MR stated in *Printing & Numerical Registering Co v Sampson* that:

> if there is one thing which more than another public policy requires it is that men of full age and competent understanding shall have the utmost liberty of contracting, and that their contracts when entered into freely and voluntarily shall be held sacred and shall be enforced by Courts of justice. Therefore, you have this paramount public policy—that you are not lightly to interfere with this freedom of contract.[19]

The free will to contract and the autonomy of parties, as discussed in Part II, are the foundations of international arbitration. The parties to an agreement who have made a choice of law to govern the underlying contract in which the arbitration agreement is contained and to which it intimately relates to are always free not only to make the choice of law but also to specify the scope of the governing law chosen by them. They are the best judges of which law best serves the interests of their contractual relationship, including the resolution of their disputes. If the parties fail to specify the limits of the chosen law by omitting to mention that it does not apply to the arbitration agreement, and without expressly giving authority to arbitral tribunals to make decisions on the validity of the arbitration agreement in accordance with some other system of law, then the chosen law for the matrix contract must apply to the arbitration agreement. The values of predictability and certainty underling the making of choice of law also raise a strong presumption in favour of a one-stop system of law to govern all aspects of the legal relationship between the parties to a matrix contract in which the arbitration agreement is contained. In those same circumstances, that freely chosen law should regulate all other aspects of the relationship of the parties because they made a choice of law for that purpose.

Professor Gary Born rightly stated in 2014 in *International Commercial Arbitration* that the:

> choice of the law applicable to an international commercial arbitration agreement is a complex subject. The topic has given rise to extensive commentary, and almost equally extensive confusion. This confusion does not comport with the ideals of international commercial arbitration, which seeks to simplify, expedite and

[19] *Printing & Numerical Registering Co v Sampson* (1875) 19 Eq 462, 465.

rationalize international dispute resolution.[20]

The same point was made 54 years earlier in 1960 by Lester Nurick that:

> [i]t is difficult enough to predict the effects of a choice-of-law clause in a domestic contract; but to do so in an international contract involves so many imponderables that it sometimes seems more like predicting the result of a lottery than a law suit.[21]

I seek to show that the complexity and confusion are largely due to the refusal or failure of such parties to appreciate and respect the legal effect and scope of the choice of law made. If the parties to international commercial disputes and arbitral tribunals keep faith with the purpose and legal effect of the applicable law chosen by the parties themselves, then the uncertainty and confusion as to the applicable law to an arbitration agreement can be reduced, if not removed. In this regard, the proposition advanced by Professor Born that an analysis of the governing law to an arbitration agreement "*begins* with the separability presumption"[22] needs qualification. This proposition has full effect if the parties have not made a choice of governing law to the matrix contract. Similarly, I argue that, on the contrary, where the parties have made a choice of applicable law to the underlying contract in which the arbitration agreement is embedded, the starting point is to ascertain the scope and full effect of the chosen law as expressed by the parties themselves and not with the separability principle. Using this approach, I provide a legal and principled basis for rethinking the purpose and legal effect of the parties making a choice of law.

II. The Foundations of International Arbitration and Choice of Law

Arbitration is consensual in nature because it is based on an agreement by the parties to resolve disputes through a third party nominated or selected by them rather than to have the dispute litigated before a court.[23] Arbitration is a private matter based on an agreement between the parties and this agreement usually covers issues such as the use of arbitration, the identity of the arbitrators and how they are appointed, the procedure to be followed and the applicable law.[24]

[20] Gary Born, *International Commercial Arbitration* (2edn, Kluwer Law International 2014) 472.

[21] Lester Nurick, 'Choice-of-Law Clauses and International Contracts' (1960) 54 Proceedings of the American Society of International Law at Its Annual Meeting (1921–1969) 56.

[22] Born (n 20) (emphasis added).

[23] New Zealand Law Commission (n 11) [98].

[24] ibid 59 [16].

Professor Henry De Vries describes arbitration as "a contractual substitute for national courts" in the title to his article because "it is a mode of resolving disputes by one or more third persons who derive their powers from agreement of the parties and whose decision is binding upon them."[25] In that sense, Professor Cindy Buys is arguably right when she states that "arbitration is all about choice."[26]

Parties to an international business transaction are free to choose the law of any state to govern their legal relationship. Party autonomy with regards to choice of law and arbitration is well established in international commercial and business transactions law.[27] Article 2 of the Hague Principles on Choice of Law in International Commercial Contracts establishes the parties' freedom to choose the law that will govern their contract.[28] Article 42(1) of the ICSID Convention requires tribunals to decide a dispute in accordance with rules of law "as may be agreed by the parties." The parties' freedom to choose the applicable law is regarded under Rome I Regulation as "the cornerstones of the system of conflict-of-law rules in matters of contractual obligations."[29] Choice of law is not only voluntary, it is also a purposeful and deliberate action. The parties voluntarily and deliberately affiliate themselves, their transaction and their dispute to the laws of a particulate state for the sole purpose of protecting their rights. Therefore, "rights are important to choice of law analysis."[30]

The internationality of a business transaction necessitates a choice of law. Where a contract or a business transaction touches two or more countries, each of which has its own substantive, procedural laws and conflict of law rules, uncertainty over the governing law to the resolution of disputes that may arise between the parties to the transaction is inevitable. Even within the domestic context in a federal country, different state laws may provide different advantages and disadvantages for the resolution of a dispute. A provision in a contract or business transaction for the applicable law and the forum for the resolution of disputes is an indispensable precondition to certainty of the law applicable to such activity. Specification of

[25] Henry P De Vries, 'International Commercial Arbitration: A Contractual Substitute For National Courts' (1982–1983) 57(1) Tulane Law Review 42, 43.

[26] Cindy G Buys, 'The Arbitrators' Duty to Respect the Parties' Choice of Law in Commercial Arbitration' (2005) 79(1) St John Law Review 59.

[27] Zhaohua Meng, 'Party Autonomy, Private Autonomy, and Freedom of Contract' (2014) 10(6) Canadian Social Science 212.

[28] Hague Conference on Private International Law, *Principles on Choice of Law in International Commercial Contracts* (The Hague: The Hague Conference on Private International Law Permanent Bureau 2015).

[29] Regulation (EC) No 593/2008 of the European Parliament and of the Council of 17 June 2008 on the law applicable to contractual obligations (Rome I) [2008] OJ L177/6, Preamble.

[30] Lea Brilmayer, 'Rights, Fairness, and Choice of Law' (1989) 98(7) The Yale Law Journal 1277, 1280.

the law applicable to the contract is a legal risk management mechanism to the extent that it helps avoid a dispute being governed by law that may be hostile to the interests of one of or even both parties, and which may be ill-suited for the nature of the dispute involved.[31] The efficient resolution of disputes depends in part on the appropriateness of the law in terms of its connection and suitability for the nature of the dispute involved. Therefore, it is imperative that the parties select a law that they consider appropriate for the resolution of their potential disputes. Professor Park states: "Discussion of future disputes when signing the contract often seems a bit like planning for divorce at a wedding feast".

Yet lack of reasonable certainty regarding the applicable norms will not usually enhance cross-border commerce, finance, or investment. While some deals may be consummated without regard to applicable law, others will not. In many contexts, multinational business enterprises will insist on calculating and balancing legal risks in making choices about their alternative commercial opportunities.

A banker may extend credit on the basis of his borrower's reputation and balance sheet. The lender will nevertheless want to know that the loan agreement, as well as any security agreement or third-party guarantee, will be enforced under the applicable law.[32]

There are, however, legal limits to the parties' autonomy to choose the law they want to govern their disputes resolution. This is because national laws and policies have implications for the recognition and enforcement of commercial and contractual agreements whether between private parties and states or just between private parties themselves. Private commercial activities take place within national territories and therefore fundamental national laws, and policies should be respected and upheld by private parties in making decisions on choice of law. Ultimately, private commercial activity will invariably be subject to various national rules and policies, at least when it comes to the enforcement of arbitral awards. Mandatory rules will include laws and policies that may apply irrespective of a choice of applicable law and the procedural regime selected by the parties.[33] For example, under Article 54(3) of the ICSID Convention, the execution of an arbitral award is governed by the laws concerning the execution of judgments in the state in which the award is sought to be enforced. By Article 25(1) of the ICSID Convention, any contracting state may notify ICSID of the class or classes of disputes which it would or would not consider submitting to the jurisdiction of ICSID. Under Article 46 of the Vienna Convention on the Law of Treaties, a state may invoke

[31] Buys (n 26) 65–66.

[32] Park (n 16) 659.

[33] Andrew Barraclough and Jeff Waincymer, 'Mandatory Rules of Law In International Commercial Arbitration' (2005) 6(2) Melbourne Journal of International Law 205.

the fact that its consent to be bound by a treaty has been expressed in violation of a provision of its internal law regarding competence to conclude treaties, thereby invalidating its consent.[34] Article III of the Convention on the Recognition and Enforcement of Foreign Arbitral Awards[35] (New York Convention) also requires each contracting state to recognise arbitral awards as binding and to enforce them "in accordance with the rules of procedure of the territory where the award is relied upon." The recognition and enforcement of the award may be refused if: the parties to an agreement to arbitrate were under some incapacity under the law applicable to them; the agreement arbitrated is not valid under the law to which the parties have subjected it; the subject matter of the difference is not capable of settlement by arbitration under the law of the country in which recognition and enforcement is sought; and if recognition or enforcement of the award is contrary to the public policy of that country.[36]

These conventions subject international arbitration to some measure of control and regulation by national laws and policies. The extent to which national law impacts choice of law decisions by a dispute resolution body depends on whether the parties have chosen national law as the governing law. Therefore, choice of law decisions by an arbitral tribunal or any dispute resolution body for that matter have to be made bearing in mind the laws of the country of choice in terms of the extent to which the subject matter is arbitrable and the actual enforceability of any ensuing award.

III. The Treatment of Choice of Law in International Arbitration

In this section, I analyse the circumstances in which arbitral tribunals have refused to enforce the parties' choice of law. A case in point where issues of choice of law and its scope and legal effect came under consideration by an arbitral tribunal was *Balkan Energy Limited v The Republic of Ghana (BELG v Ghana)*.[37] A close reading of the arbitration clause under consideration and the fact that the parties had agreed for the underlying contract, a Power Purchase Agreement (PPA), to be the *entire agreement* reveal that a dispute between the investor and Ghana about the law governing the validity of the arbitration clause was supposed to have been

[34] Vienna Convention on the Law of Treaties (adopted 23 May 1969, entered into force 27 January 1980) 1115 UNTS 331 < https://treaties.un.org/doc/publication/unts/volume%201155/volume-1155-i-18232-english.pdf>.

[35] Convention on the Recognition and Enforcement of Foreign Arbitral Awards (adopted 10 June 1958, entered into force 07 June 1959 <https://www.uncitral.org/pdf/english/texts/arbitration/NY-conv/New-York-Convention-E.pdf>.

[36] ibid article V.

[37] *Balkan Energy Limited (Ghana) v The Republic of Ghana*, PCA Case No 2010–7, Interim Award, 22 December 2010.

resolved in terms of the governing law to the main contract and the arbitration rules chosen by the parties in the underlying contract. An in-depth discussion of this case is necessary to establish the point that the purpose and legal effect of the parties making a choice of law were not appreciated in this case. It would also seem that the Tribunal sought to avoid the application of the chosen law to assert its own jurisdiction and ultimately make a decision that favours the private party to the suit.

Balkan Energy Ltd (Ghana) (BELG) was a limited liability company incorporated under the laws of Ghana. Its sole shareholder was Balkan Energy Limited (incorporated in the United Kingdom), which was in turn wholly owned by Balkan Energy LLC of the United States. Balkan Energy Ltd entered into a PPA with Ghana in 2007 for the refurbishment and commissioning of dual fired (diesel and gas) barge and associated facilities. The PPA provided "if *any* dispute *arises out of or in relation to this Agreement* and if such matter cannot be settled through direct discussion of the Parties, the matter shall be referred to binding arbitration… Arbitration shall be governed by and conducted in accordance with UNCITRAL rules".[38] Thus the arbitration agreement was contained in the matrix contract (the PPA), which itself was governed by and was to be construed in accordance with the laws of Ghana.

Balkan Energy Ltd alleged a breach of the PPA and commenced arbitration proceedings against Ghana.[39] While the terms of appointment of arbitrators confirmed the UNCITRAL Rules as the governing rules of the arbitration and the laws of Ghana as the governing law of the PPA,[40] the parties disagreed as to which law governed the arbitration agreement. Dutch law (*lex loci arbitri*) was favoured by BELG, while the State maintained that Ghanaian law was the proper law.[41] The State argued that Ghanaian law governed the arbitration clause and also determined issues of arbitrability because the parties "specifically subjected the PPA to the laws of Ghana, and the default position in international arbitration is that a choice of law provision in the main contract also applies to an arbitration clause contained therein."[42] The BELG argued that the "conflict of laws rules… should not be used to determine the law applicable to arbitrability, but, rather, only to determine the substantive contract law applicable to the arbitration

[38] *Power Purchase Agreement between the Government of Ghana Acting through its Minister for Energy and Balkan Energy (Ghana) Ltd on Osagyefo Power Barge and Associated Facilities*, signed on 27 July 2007, clause 22.2 (emphasis added).

[39] *BELG v Ghana* (n 37) [4], [6]–[7].

[40] ibid [17].

[41] ibid [98].

[42] ibid [67].

agreement."[43] It argued that even if the formal validity of the arbitration clause was governed by Ghanaian substantive contract law, Ghanaian law did not govern issues of arbitrability because tribunals "usually determine the arbitrability of a dispute on the basis of the law of the place of arbitration, except in exceptional cases involving matters of public policy."[44]

Ghana objected to the jurisdiction of the Interim Tribunal, arguing that both the PPA and the arbitration clause were void because the PPA did not receive Parliamentary approval as required by Article 181(5) of the Constitution of the Republic of Ghana 1992 (the Constitution).[45] Ghana maintained that both the PPA and the arbitration clause were an "international business or economic transaction," and were therefore void and unenforceable due to lack of prior parliamentary approval.[46] Ghana argued that the determination of the validity of either the PPA or the arbitration clause involved questions of interpretation of the Constitution and was, therefore, non-arbitrable.[47]

The Tribunal held that it was competent to decide on the validity of the arbitration agreement and had the jurisdiction to entertain the substantive suit under the competence-competence and separability principles in international arbitration which allow tribunals to determine questions concerning their jurisdiction and the substantive validity of the arbitration clause respectively.[48] According to the Tribunal, the arbitration clause was valid because all the formalities for its validity had been met.[49] The Tribunal held that the law applicable to the arbitration agreement was the law of the seat of arbitration—namely, The Netherlands.[50] It reasoned that:

> The constitutional interpretation issue did not necessarily fall within the public policy restriction in The Netherlands, because, if it did, any arbitration agreement or contract which encounters an argument of constitutional nature in a foreign country would be excluded from the jurisdiction of an arbitration tribunal to decide

[43] ibid [88].
[44] ibid.
[45] ibid [9].
[46] ibid [117]–[119].
[47] ibid [64].
[48] ibid [100], [115], [152]–[153], [167] and [99].
[49] ibid [147].
[50] ibid [152]–[153].

upon.[51]

According to the Tribunal, the issue of the substantive validity of the PPA under the Constitution of Ghana and its consequences for the rights and obligations of the parties pursuant to their contractual undertakings were, "in essence, questions pertaining to the merits of the dispute, and they do not affect the validity or otherwise of the arbitration agreement."[52] The Tribunal held that there were strong arguments to be made:

> in favour of defining the scope of arbitrable matters in accordance with the *lex loci arbitri*… [T]he Parties' agreement to dispute settlement before the PCA is an indicator that the Parties intended to remove questions relating to dispute resolution—as opposed to the substantive performance of the contract—from the place of either Party, to a neutral forum.[53]

The Tribunal also saw "no reason, under Dutch or Ghanaian law, that constitutional provisions should be inherently non-arbitrable."[54] The position of the Tribunal seems to suggest that the law of the place of arbitration is not only mandatory and automatically applicable to the dispute but that it is inherently neutral because of the choice of the place as the seat of arbitration. This is quite misleading. It is arbitration that was chosen as a 'neutral' mechanism for the settlement of the parties' disputes and not the laws of The Netherlands. The Netherlands was the seat of arbitration because it happened to have been chosen by the parties, perhaps given the facilities it has for such arbitration. As argued by Professor Jan Paulsson, any other place could have been chosen as the seat of arbitration "[u]nless there are objective reasons to conclude that a *situs* is hostile to awards rendered in compliance with the Rules agreed between the parties, it is assumed that the whole world is a possible *situs*."[55]

Professor Gary Born states that an "[a]nalysis of choice of law governing an arbitration agreement begins with the separability presumption" by which an international arbitration agreement "is presumably separable from the underlying contract with which it is associated."[56] The effect of this presumption is that the arbitration agreement may be governed by a different law than the law governing

[51] ibid [144].
[52] ibid [112].
[53] ibid [142].
[54] ibid.
[55] Jan Paulsson, 'Delocalisation Of International Commercial Arbitration: When and Why It Matters' (1983) 32(1) The International and Comparative Law Quarterly 53, 55.
[56] Born (n 20) 472.

the primary contract.[57] In the words of Professor Born, the "separability doctrine does not mean that the law applicable to the arbitration clause is *necessarily* different from that applicable to the underlying contract. It instead means that differing laws *may* apply to the main contract and the arbitration agreement."[58] As such, it is not conclusive or settled that the choice of a place as the seat of arbitration means the laws of the seat of arbitration become automatically applicable to the arbitration or an issue arising out of that arbitration. Indeed, the Inter-American Convention on the Law Applicable to International Contracts directly addresses this issue by stating in Article 7 that the "[s]election of a certain forum by the parties does not necessarily entail selection of the applicable law."[59] Professor Born also points out that "different laws may apply to issues of formal validity, substantive validity, capacity, interpretation, assignment and waiver of an international arbitration agreement."[60] It follows that, in the instant case, the fact that The Netherlands was the seat of arbitration did not mean that its laws enjoyed some particular advantage or were inherently neutral and necessarily became automatically applicable to the parties' dispute as to the validity of the arbitration clause.

The application of Dutch law would invariably result in a decision being made in favour of one of the parties just as it would have been the case if the governing law was applied. The law of the seat of arbitration does not enjoy any particular advantage over the chosen law in terms of neutrality when it comes to its practical application to the dispute, especially when the governing law was not enacted specifically to give unfair protections to one of the parties to the detriment of the other party. Thus, if The Netherlands as the seat of arbitration was neutral, it would be neutral because it was *the place of arbitration* and, more importantly, for its presumed neutrality of those who sit to arbitrate. This, however, did not automatically and inherently import neutrality in the practical application of Dutch law. As Professor Lea Brilmayer rightly states:

> [w]henever a law is applied it will work to the advantage of one party to the litigation and to the disadvantage of the other. Choice of law at the adjudicative stage is a zero sum game; what advances the cause of the plaintiff simultaneously imposes costs on the defendant, and vice versa. This is as true in purely domestic cases as in conflicts cases.[61]

[57] ibid.

[58] ibid 475 (emphasis original).

[59] Inter-American Convention on The Law Applicable to International Contracts (signed 17 March 1994) (Fifth Inter-American Specialized Conference on Private International Law), signed at Mexico, DF Mexico.

[60] Born (n 20) 472.

[61] Brilmayer (n 30) 1280.

Thus, if by the text of the parties' agreement they meant to have the chosen law to govern the validity of the contract and the arbitration agreement, then this must be so upheld and respected whatever the outcome.

Professor Born points out that "by applying a law other than that governing the parties' underlying contract, national and international tribunals have sought to safeguard international arbitration agreements against challenges to their validity based on local (often idiosyncratic or discriminatory) law."[62] This, it is said, promotes the enforceability of arbitration agreements, clauses, and the general efficacy of arbitration.[63] Tribunals must not avoid the application of the choice of law made by the parties to an arbitration agreement merely because by applying that choice of law the arbitration agreement will be invalidated. Effect must be given to the intent of the parties to arbitrate and an agreement to arbitrate must not be invalidated lightly. If, however, by express requirement or by necessary and reasonable implication the governing law to the underlying contract applies to the arbitration agreement, it must be so held even if the arbitration clause will thereby be invalidated. If an implication can legitimately be made from the parties' agreement that the governing law to the contract shall apply to the arbitration agreement, that will be consistent with the intent of parties. And if the parties do not give a tribunal the authority to make decisions out of its sense of equity and of right, ignoring this implication deriving from the parties' agreement substitutes the tribunal's will and wish for the intent of the parties. Arbitral tribunals must not have interest in upholding the validity of an arbitration agreement beyond the express and implied understanding of the parties. Moreover, a refusal to apply the chosen law to both the underlying contract and to the arbitration agreement when it does apply expressly or by implication defeats the quest for *certainty and predictability* as to the applicable law which underlies the parties' decision to make a choice of law in the first place. After all, simply following the separability principle does not necessarily always promote the course of the arbitral process. As Professor Born states:

> An unfortunate consequence of the separability presumption in the choice-of-law context has been the development of a multiplicity of different approaches to choosing the law governing the formation, validity and termination of international arbitration agreements. National courts, arbitral tribunals and commentators have adopted a wide variety of choice-of-law approaches to issues of substantive validity, ranging from application of the law of the

[62] Born (n 20) 475.
[63] ibid.

> judicial enforcement forum, to the law of the arbitral seat, to the law governing the underlying contract, to a "closest connection" or "most significant relation" standard, to a "cumulative" approach looking to the law of all possibly-relevant states.

Other authorities have suggested even more esoteric choice-of-law rules, including the law of the arbitrator's residence or *lex mercatoria*. Commentators have variously identified three, four, or as many as nine approaches to the choice of law governing international arbitration agreements.

This multiplicity of choice-of-law rules potentially applicable to the arbitration agreement does not advance the purposes of the international arbitral process. The existence of multiple choice-of-law rules creates unfortunate uncertainties about the substantive law applicable to arbitration agreements, as well as the risk of inconsistent results in different forums.

In turn, this leads to uncertainty about the extent to which international arbitration agreements can actually be relied upon to provide an effective means of resolving international disputes. The multiplicity of choice-of-law rules also leads to delays and expense, resulting from the need to engage in choice-of-law debates, before both arbitral tribunals and national courts, when disputes arise concerning the formation or validity of arbitration agreements. This is inconsistent with parties' expectations of an efficient, centralized dispute resolution mechanism in entering into international arbitration agreements.[64]

In this regard, Professor Born proposes that "the analytical confusion about choice-of-law questions regarding the arbitration agreement creates uncertainty, delay and the risk of inappropriate and unjust results, and should be clarified."[65] The parties' choice of law was intended to clarify this uncertainty and avoid unnecessary delays and costs when the parties have made choice of law. Arbitral tribunals should not seek to sever or detach the law of the host country (which has also been chosen by the parties as the applicable law) for purposes of establishing their jurisdiction or validating an agreement to arbitrate and at the same time expect the investor to rely on these same laws for the purposes of enforcing an ensuing award. If indeed the choice of a place as the seat of arbitration is meant to completely detach the laws of the countries of the parties from the dispute, then the winning party cannot rely on the laws of the host country to enforce an award arising from the settlement of the dispute. The parties to an arbitration agreement are at liberty to choose the law of the seat of arbitration. If they do not do so, it

[64] ibid 476–477.
[65] ibid 46.

should not be lightly presumed that just because they chose the place as the seat of arbitration, they meant for the law of the seat to apply to their dispute.

On the issue of the applicable law to the arbitration clause itself, in the *BELG v Ghana* case, Ghana argued that since the PPA was governed by the laws of Ghana, a matter which was not disputed by BELG, the laws of Ghana should also apply to the arbitration agreement contained in the PPA. Balkan Energy Ltd argued to the contrary that since the PPA "did not make an express choice of law in respect of the arbitration agreement," the arbitration agreement was governed by the law of the seat of arbitration, in this case Dutch law.[66] The parties had specified that the governing law to the PPA was Ghanaian law. The "arbitration" was to be governed by and construed in accordance with UNCITRAL Arbitration Rules.[67] The parties did not expressly, positively or directly state the law to govern the arbitration clause. It is, however, clear that they expressly stated that *any* dispute which arose out of or related to the PPA had to be resolved in accordance with UNCITRAL Arbitration Rules. Since the dispute about the validity of the arbitration clause arose out of and in relation to the PPA, the parties must have intended for the laws governing law the PPA and the arbitration rules governing the arbitration to apply to the arbitration clause as well. Nonetheless, the Tribunal chose to apply the separability principle and concluded that the law of the seat of arbitration applied.

The parties had agreed to the applicable law to the arbitration, namely UNCITRAL Arbitration Rules for purposes of predictability and certainty. Since the arbitration was the mode of resolving the dispute and it was agreed to by the parties through an arbitration clause, both of which manifestly related to the PPA, the governing law to the arbitration (UNCITRAL Arbitration Rules) and the governing law to the main contract (Ghanaian Law) applied to the arbitration clause and took precedence over the law of the seat of arbitration. The governing law to the arbitration and the law to the main contract were manifestly closest to the arbitration clause than the law of the seat of arbitration. In particular, the law governing the arbitration as specified by the parties themselves was closest to the arbitration clause since the arbitration arose only because the parties agreed to submit the dispute to arbitration. In *Tamil Nadu Electricity Board v ST-CMS Electric Company Private Ltd*, it was held that the parties "have agreed to arbitration in accordance with English law and it is by that law alone that the ambit of the

[66] *BELG v Ghana* (n 37) [148].

[67] UNCITRAL Rules on Transparency in Treaty-based Investor-State Arbitration "UNCITRAL Arbitration Rules (with new article 1, paragraph 4, as adopted in 2013)" (Vienna: UNCITRAL February 2014) <http://www.uncitral.org/pdf/english/texts/arbitration/arb-rules-2013/UNCITRAL-Arbitration-Rules-2013-e.pdf>.

arbitration provision can be determined, as a matter of construction."[68] Ghana and BELG had agreed to arbitration in accordance with UNCITRAL Arbitration Rules. Moreover, between the governing law to the main contract and the law of the seat of arbitration, the former was closer to the arbitration clause since the substantive dispute and the dispute about the validity of the arbitration clause related to or arose directly out of the underlying contract. Thus, the law of the seat of arbitration was comparatively far removed from the main contract and the dispute arising out of the performance of that contract and should have come last in the consideration of the applicable law to the arbitration clause. The parties' choice of the governing law *to the arbitration* and the governing law to the main contract legally justifies such an approach argued for here. Article 35(3) of UNCITRAL Arbitration Rules states that *"[i]n all cases*, the arbitral tribunal shall decide in accordance with the terms of the contract, if any, and shall take into account any usage of trade applicable to the transaction".[69] "In all cases" in this provision would include situations where the validity of the arbitration clause is in dispute. The "terms of the contract" include the governing law to arbitration and the governing law to the main contract; these ought to have been applied in determining the validity of the arbitration clause.

The Tribunal's application of the law of the seat of arbitration could only be justified if the parties did not make a choice of applicable law to both the arbitration and the underlying contract. The decision of the Tribunal shows that it was out to make a decision that gave effect to the arbitration agreement, irrespective of the legal effect of the parties making a choice of law without limiting the scope of application of the governing law they have chosen. This is because the position of the Tribunal was very linear and lopsided towards validating the arbitration agreement rather than objectively deciding on the issues. In other words, the Tribunal started from an approach that was biased in favour of justifying or validating the agreement to arbitrate and favouring the rights of BELG rather than to assess the merits of each claim objectively from a neutral position. Holding in favour of BELG on this issue, The Tribunal stated:

> in deciding this issue, it should favour the approach that is more conducive to making the arbitration agreement effective rather than an approach that would render the agreement ineffective. The Parties agreed to an arbitration clause providing for the resolution of disputes arising under the PPA by arbitration and it is this choice that should prevail and not an interpretation the

[68] *Tamil Nadu Electricity Board v ST-CMS Electric Company Private Ltd* [2007] EWHC 1713, [2007] 2 All ER (Comm) 701, [35].

[69] UNCITRAL Arbitration Rules (n 67) (emphasis added).

> result of which would be the exact opposite. A contract cannot be deemed to contain a clause which is self-defeating of its objectives. The validation principle invoked by the Claimant lends support to the conclusion that it makes more sense to consider that the Parties opted for an approach that would validate rather than render invalid the arbitration agreement.

The solution to this issue is also not clear-cut by reference to conflict of law rules. The basic tenet underlying the doctrine of *lis pendens*, however, points in the direction of finding in favour of the law that is most closely connected to the arbitration agreement. In this case it is the law of the Netherlands that appears to have the closest connection with the arbitration agreement under the PPA. This is borne out by the fact that The Netherlands was chosen as the seat of the arbitration and by the explicit decision to operate under the UNCITRAL Rules, which, among other consequences, determines the courts which will be competent to consider any challenge to the award rendered. More important still is the argument invoked by the Claimant to the effect that the choice of the seat of the arbitration in a neutral country indicates a clear understanding that the Parties wish to detach the arbitration agreement from the domestic law or the courts of either Party. The situation is, of course, different with respect to the law applicable to the PPA, since the PPA contains a choice of law provision that expressly subjects the contract to Ghanaian law.

In the light of the above considerations, the Tribunal concludes that the law applicable to the arbitration agreement in the PPA is the law of The Netherlands. In so deciding, the Tribunal wishes to state that this entails no disrespect for the laws of Ghana or of other developing countries. The Tribunal is sensitive to the importance of according due respect to the laws of every sovereign State; and it emphasizes that its decision in the present case is entirely unrelated to any views or judgments regarding the merits of the respective legal systems. Rather, its decision is based solely on its appreciation of which solution appears to be more appropriate for the effective discharge of the dispute resolution functions which have been entrusted to it by the agreement of the Parties themselves.[70]

The Tribunal ultimately concluded that the proper law governing the validity of the arbitration agreement was Dutch law and that therefore Article 181(5) of the Constitution did not in any way affect the validity of the arbitration agreement.[71] For the Tribunal then, the "more appropriate" solution for the "effective"[72] discharge of the dispute resolution functions entrusted to it was to

[70] *BELG v Ghana* (n 37) [149-150] and [152].
[71] ibid [154].
[72] ibid [152].

make a finding that the arbitration agreement is valid and binding. The position of the Tribunal is very problematic because it failed to explain why the same parties will make a choice of law of a place other than that of the seat of arbitration to govern the subject matter of the contract if indeed "the choice of the seat of the arbitration in a neutral country indicates a clear understanding that the Parties wish to detach the arbitration agreement from the domestic law or the courts of either Party."[73] The Tribunal did not give reasons for why a finding of the invalidity of the arbitration agreement under the governing law of the PPA chosen by the parties themselves was not at all appropriate and legitimate, or at least less so in comparison with its finding that the arbitration was binding under Dutch law. It is as if an arbitration agreement must always be valid and binding. The conflict of laws rule requiring the application of the law of the seat of arbitration is a default rule to comply with only if the parties have not made a choice of law that expressly or impliedly applies to the arbitration clause. If the choice of the seat of arbitration means the law of the seat automatically governs the agreement to arbitrate because of its *neutrality*, such logic should extend to the application of the law of the seat to the subject matter of the contract because the chosen law cannot truly be said to be neutral since it is the law of one of the parties. If this logic about the 'neutrality' of the law of the seat of arbitration was to be adopted *simpliciter*, then making a choice of law becomes of no use.

Indeed, the parties to the PPA recognised that any of its provisions could be held to be invalid when they agreed in clause 25 that "this Agreement shall be construed, if possible, in a manner to give effect by means of valid provisions to the intent of the parties to the particular provision or provisions held to be invalid." This provision specifically referred to the possibility of a court of competent jurisdiction holding the provisions of the PPA to be invalid. Similarly, in clause 27, the parties agreed that the PPA "contains all of the understandings and agreements of whatsoever kind and nature with respect to the subject matter of this Agreement and the rights, interests, understandings, agreements and obligations of the parties relating thereto." The parties made express reference to the laws of Ghana and the arbitration rules in clauses 22 and 23 of the PPA to govern the PPA and the arbitration. Based on clause 27 of the PPA, their rights and obligations "with respect to the subject matter" (the PPA) could not be determined outside of the scope of the PPA and the governing law they had expressly incorporated into the PPA. Since the PPA was the *entire agreement*, the "understandings" of the parties according to clause 27 was that the arbitration clause which was embedded in the PPA was to be governed by the law chosen to govern the PPA just like any other term of the PPA was. Thus, it was only if the parties did not agree on a choice

[73] ibid [149].

of law to govern the PPA and the law governing the arbitration that the Tribunal could embark on a journey in search of a law that would allow it to give effect to the arbitration agreement.

The claimant had argued that the parties had "not expressly agreed on the governing law" to the arbitration agreement and that in "such situations, it is common for arbitral tribunals to refuse to apply the general choice-of-law clause in the main contract to the arbitration agreement, *especially where the Parties' chosen law would invalidate the arbitration clause*."[74] Clearly, the concern of the claimant was not that the chosen law had no legitimate application to the arbitration agreement; its primary concern was that the application of the chosen law would invalidate the arbitration agreement. Thus, this was a case of a party seeking to avoid the consequences of applying the chosen law it had voluntarily agreed to. This is certainly not a legitimate basis to refuse to apply the chosen law, especially when the PPA and laws expressly stated in it as the governing laws constituted the entire legal regime for their rights and obligations. The Tribunal emphasized that the New York Convention "supports the conclusion that, in the absence of a choice of law provision in the arbitration agreement, the law of the seat of arbitration should be the applicable law for determining the validity of the arbitration agreement."[75] This position can only hold if the arbitration agreement is not embedded in the matrix contract. The New York Convention does not say that the governing law to a substantive contract cannot affect an agreement contained in it. The Convention also does not say that the parties must agree to a separate choice of law clause to an agreement to arbitrate contained in the contract. Therefore, the parties may well agree that the governing law to the contract covers an agreement to arbitrate and that this can be express or implied. In this case, the text of the arbitration clauses showed that the governing law to the arbitration was to be applied in determining the validity of the clause.

The conclusion of the Tribunal that "the arbitration agreement embodied in… the PPA is both valid and enforceable *independently from the issue of the validity of the PPA*"[76] was inaccurate and misleading as to what the real issue was. The real issue was whether the arbitration agreement "embodied in… the PPA" was valid and enforceable in light of the governing law to the PPA in which it was embodied. It was not whether the validity of the arbitration agreement was dependent on the validity of the PPA. The fact that both could be invalidated on the basis of the

[74] ibid [131] (emphasis added).
[75] ibid [151].
[76] ibid [167] (emphasis added).

governing law does not mean that the validity of one was dependent on the validity of the other.

Article 23 of UNCITRAL Arbitration Rules states that for the purpose of giving effect to the power of a tribunal to rule on its own jurisdiction:

> an arbitration clause that forms part of a contract shall be treated as an agreement independent of the other terms of the contract. A decision by the arbitral tribunal that the contract is null shall not entail *automatically* the invalidity of the arbitration clause.[77]

This provision does not say that an agreement to arbitrate can *never* be invalidated based on the invalidity of the underlying contract. It also does not say that the parties cannot agree to a choice of law that covers both the underlying contract and its agreement to arbitrate. If the parties can choose a law to govern both the subject matter of the contract and the arbitration agreement, then that chosen law can determine the validity of both. Thus, the requirement that an agreement to arbitrate be treated as *independent of the contract* does not mean that the governing law cannot cover and affect the agreement to arbitrate if by specifying the governing law the parties expressly or impliedly intended that it should cover and affect all other terms of the contract. Indeed, the courts are willing to hold that an arbitration *contained in the matrix contract* is invalid and not binding if one of the parties does not sign the matrix contract.[78]

There is no legal requirement that the parties agree to separate governing laws—one to the matrix contract and the other to the arbitration agreement. Thus, if the parties do not expressly exclude an agreement to arbitrate from being covered by the governing law to the matrix contract in which the arbitration agreement is embodied, then the law chosen by the parties governs the arbitration agreement as a term, like all other terms of the matrix contract. As Dr Ronán Feehily rightly argues, since the arbitration clause and the main contract "are included in one document signed by the parties, the separability doctrine has been characterised as solely a legal concept and not a factual determination… [T]here is no requirement on the parties to separately consent to the arbitration agreement."[79] Thus, the object of Article 23 of UNCITRAL Arbitration Rules is to preserve the jurisdiction of the arbitral tribunal, so that the tribunal can at least commence proceedings to

[77] UNCITRAL Arbitration Rules (n 67) (emphasis added).

[78] *Fiona Trust and Holding Corporation and Others v Yuri Privalov and Others* [2007] UKHL 40, [2007] 4 All ER 951.

[79] Ronán Feehily, 'Separability in International Commercial Arbitration; Confluence, Conflict and the Appropriate Limitations in the Development and Application of the Doctrine' (2018) 34 Arbitration International 355, 356.

decide on the validity of an agreement to arbitrate in the first place. The object of the provision is not to say that the validity of an agreement to arbitrate cannot be questioned. The arbitration agreement, Professor Dietmar Czernich argues, "is the bridge from the land of litigation to the land of arbitration. This bridge has to hold so that the litigants can reach the land of arbitration."[80] I argue that the fate of such an agreement should depend on the intention of the parties to the matrix contract and arbitration agreement as revealed in their choice of law. The values of predictability and certainty of the applicable law underlying the making of choice of law strongly support the presumption of a one-stop legal regime to regulate all aspects of the parties' legal relationship.

In *BELG v Ghana*, the Tribunal maintained that Ghana was "attempting to rely on its own domestic law to invalidate an arbitration agreement to which it previously acceded and implemented, an approach that Dutch law and international law and international public policy do not permit."[81] The issue with this proposition is that it is lopsided. First, the crucial thing is that Ghanaian law was the governing law voluntarily chosen by the parties. Based on the parties' choice, either party was entitled to rely on the chosen law in support of their legal argument if it worked in their favour. That was exactly what Ghana did. Ghana did not make the simple argument that its law must invalidate the agreement to arbitrate. That argument was based on the fact the parties, who were at liberty to choose the law of the seat of arbitration, chose Ghanaian law as the governing law. Second, given that the parties had chosen Ghanaian law, which is domestic law, it is not convincing that Dutch law (which is also domestic law) should dictate whether Ghanaian law applied or not. The fact that Dutch law was *lex loci arbitri* did not remove the fact that the parties had made a choice of law which entitled a party to rely on that choice of law. Third, the tribunal failed to articulate the content of the so-called "international law and international public policy" and how they denied Ghana from arguing based on its law which was also chosen by the parties. The Tribunal also held that "reliance on internal law should not be permissible to invalidate an arbitration agreement whether the place of performance is within or outside of the State."[82] This argument could also hold if the parties did not choose Ghanaian law. If the parties choose internal law and that internal law governs both

[80] Dietmar Czernich, 'The Theory of Separability in Austrian Arbitration Law: Is It on Stable Pillars?' (2018) 34 *Arbitration International* 463.

[81] *BELG v Ghana* (n 37) [160].

[82] ibid [161].

the substantive contract and the arbitration agreement, then the position held by the tribunal will lack merit.

The Tribunal also held that "[w]hile there may be valid reasons for the Ghanaian Constitution to impose restrictions on the State's powers to enter into certain kinds of international transactions, there are circumstances where such a restriction cannot derogate from the effectiveness of the arbitration agreement to which the Parties are committed and which has been held out as valid by the competent Ghanaian officials."[83] It should also be noted the fact that Ghana "proposed the alternative of arbitration to settle the dispute indicates that the arbitration agreement was considered by it to be valid and in force, even if" Ghana "ultimately decided not to pursue this line of action."[84] For the Tribunal, "the approval of business transactions by government officials not objected as to their legality under local law, and relied upon for a number of years, have been held to amount to estoppel by various arbitral tribunals."[85]

The efficacy of these propositions depends on whether the chosen law of the parties governed the substantive contract and the arbitration agreement. If the chosen law which included the Constitution governed the arbitration agreement, the validity of that agreement could be questioned within the terms of the chosen law. It is neither consistent with the purpose of making a choice of law nor fair to avoid the application of the chosen law simply because its application would invalidate an agreement to arbitrate. The parties to a choice of law clause in making that choice of law must accept the natural and inherent consequences attached to the application of their choice of law. Thus, if parties make a choice law without *expressly* making room for justice to be done as between them in accordance with any other law, then matters and disputes arising between them must be resolved in accordance with the chosen law. In this regard, the choice of a location as the place of arbitration is for that location to play the role of the seat of arbitration. If the parties intended the law of the seat of arbitration rather than the governing law to regulate the conduct of the arbitration in any particular manner, they would have expressly agreed to that because it was their free will to do so. Thus, the choice of a place as the seat of arbitration alone without more cannot make the *lex loci arbitri* trump the governing law freely chosen by the parties if the parties did not limit the scope of application of the governing law.

It is unusual that if the arbitration clause is invalid under the law chosen to govern the contract but valid under the law of the seat, then some tribunals and commentators may prefer the law of the seat to be applied to the arbitration

[83] ibid [163].
[84] ibid.
[85] ibid [165].

clause based on the separability principle. The separability presumption, however, provides a rebuttable basis for the law of the seat of arbitration to be considered to uphold the validity of the clause. Where the parties have made a choice of law, the separability presumption does not provide legal justification as to why the arbitration clause must not be held invalid under the governing law that expressly or impliedly applies to that clause. When the parties have made a choice of law, it is the intent of the parties as revealed in the express terms of the choice of law clause or as implied from the purpose of making the choice of law that provides that legal justification to determine the validity of the arbitration clause based on that chosen law. In other words, the express or implied intent of the parties as revealed in the making of the choice of governing law and deduced from the text of the matrix contract read as a whole supersedes the rebuttable presumption of separability. In the 1894 English case of *Hamlyn & Co v Talisker Distillery*, it was held that the governing law depends on the parties' intent "either as expressed in their contract, or as derivable by fair implication from its terms."[86]

In fact, recent case law does not limit the proper law of the arbitration clause to the law of the seat of arbitration even if the parties have not agreed to the governing law to the arbitration agreement. The English common law system, for example, recognises that the law applicable to the arbitration clause may be determined by taking into consideration a three stage approach: whether the parties have expressly chosen the governing law of the arbitration clause; whether the parties have *impliedly* made a choice of the governing law to the arbitration clause; and in the absence of an express or implied choice of law, a determination is made as to the law which the arbitration agreement has the closest, most real or substantial connection to. This is reflected in English cases such as *Sulamérica Cia Nacional De Seguros SA v Enesa Engenharia SA*,[87] *Arsanovia Ltd v Cruz City 1 Mauritius Holdings*,[88] *Habas Sinai Ve Tibbi Gazlar Istihsal Endustrisi AS v VSC Steel Company Ltd*,[89] *Fiona Trust and Holding Corporation v Yuri Privalov*[90] and *Tamil Nadu Electricity Board v ST-*

[86] *Hamlyn & Co v Talisker Distillery* [1894] AC 202 (HL) 212.

[87] *Sulamérica Cia Nacional De Seguros SA v Enesa Engenharia SA* [2012] EWCA Civ 638, [2013] 1 WLR 102.

[88] *Arsanovia Ltd v Cruz City 1 Mauritius Holdings* [2012] EWHC 3702 (Comm), [2013] 2 All ER (Comm) 1.

[89] *Habas Sinai Ve Tibbi Gazlar Istihsal Endustrisi AS v VSC Steel Company Ltd* [2013] EWHC 4071 (Comm), [2013] 2 All ER (Comm) 1.

[90] *Fiona Trust* (n 78).

CMS Electric Company Private Ltd.[91] In the United States, the courts have also implied intent in the matter of the governing law.[92]

These cases support the proposition that, in conceptual terms, an *express* choice of law in the matrix contract is an *implied* choice of the law governing the arbitration clause.[93] It might be argued that it makes no sense to make such an implication if the arbitration clause is invalid under the law chosen to govern the matrix contract but valid under the law of the seat. The desire, however, to uphold the validity of the arbitration agreement using the separability principle does not provide a more solid legal justification to apply the law of the seat of arbitration rather than to imply the application of the chosen law to the matrix contract to the arbitration agreement when such implication can be derived from the intent of the parties to the arbitration agreement (as can be revealed from their choice of law). As I argued earlier, if the parties who are at liberty to choose the law of the seat of arbitration as the governing law do not do so, it is inconsistent with party autonomy, which is the foundation of arbitration, to say that because the parties have chosen a place as the seat of arbitration, they thereby implied that the law of the seat should apply to the arbitration clause.

It is in accordance with the parties' desire for predictability and certainty as to the applicable law in making a choice of the governing law that that the law should apply in determining the validity of the arbitration agreement. Since arbitration, a product of the parties' agreement to arbitrate, is about a dispute that arises out of the performance of the main contract, there is a manifest, substantial, closest, most real or significant connection between the substantive contract and the dispute and between the arbitration clause and the matrix contract. The chosen law must therefore govern the matrix contract and the arbitration agreement. The application of the law of the seat of arbitration becomes relevant for consideration just like any other connecting factor where no express choice of law is made. In the case of Member States of the European Community, Rome I Regulation states that:

> [w]here all the elements relevant to the situation at the time of the choice are located in a country other than the country whose law

[91] *Tamil Nadu Electricity Board* (n 68).

[92] *Connecticut General Life Insurance Co v Boseman*, 84 F (2d) 701 (CCA 5th, 1936); *Boseman v Connecticut General Life Insurance Co* 301 US 196 (1937); *Liberty National Bank and Trust Co. v. New England Investors Shares*, 25 F (2d) 493 (D Mass 1928); *Russell v Pierce*, 121 Mich 208, 80 NW 118 (1899); *United States Savings & Loan Co v Shain*, 8 ND 136, 77 NW 1006 (1898); and *Midland Savings & Loan Co v Henderson*, 47 Okla 693, 150 Pac 868 (1915). See also Symeon Symeonides, 'Choice of Law in the American Courts in 1997' (1998) 46(2) The American Journal of Comparative Law 233; and Gregory E Smith, 'Choice of Law in the United States' (1986–1987) 38 Hastings Law Journal 1041.

[93] Born (n 20) 524.

> has been chosen, the choice of the parties shall not prejudice the application of provisions of the law of that other country *which cannot be derogated from by agreement*".[94]

The addition of the emphasized phrase suggests that it is not sufficient to apply the law of a country other than that of the chosen country merely because the elements are located in that other country. Rather, it has to be shown that the law in that other country cannot be derogated from by agreement.

IV. The Obligation to Apply the Chosen Law

The *BELG v Ghana* case considered above highlights the important issues raised by Daniel Hochstrasser—namely, "whether the choice of law made by the parties limits the arbitral tribunal with respect to the applicable law to such provisions which are part of the legal system designated by the parties as the *lex contractus*."[95] Addressing this issue requires a language analysis of the provision expressing the choice of the governing law to the contract, the text of the arbitration clause, and the provisions of conventions that govern the exercise of powers by an arbitral tribunal. The analysis below shows that international arbitration conventions require arbitral tribunals to keep faith with the parties' choice of law. The scope of the chosen law and the limitations of choice of law on the powers of arbitral tribunals over whether to depart from that chosen law have not been considered in arbitral proceedings; it was certainly not argued in *BELG v Ghana*. I argue that a choice of law made by the parties limits the competence of a tribunal in determining the applicable law to the express and implied laws of the country designated by the parties, including its rules of private international law. The ICSID Convention states in Article 42(1) that a tribunal:

> shall decide a dispute in accordance with such rules of law as may be agreed by the parties. In the absence of such agreement, the Tribunal shall apply the law of the Contracting State party to the dispute (including its rules on the conflict of laws) and such rules of international law as may be applicable.

By this provision, the primary obligation of a tribunal when there is a choice of law is to apply the law chosen by the parties. This provision prevents a tribunal from fishing for some other law when the parties have legitimately made a choice of law. This is the case even if the consequences of applying the choice

[94] Council Regulation (EC) 593/2008 (n 29) article 3(3).
[95] Hochstrasser (n 18) 57.

of law will be unpalatable to a tribunal, both the parties, or one of them. The provision does not make room for a tribunal to apply some other law when the parties have made a choice of law. It is only in the absence of such agreement on choice of law that a tribunal is bound to apply the law of a *contracting state party* to the dispute. The law of the contracting state in such circumstances is defined to include the state's rules on the conflict of laws. If the parties make a choice of law, rules of international law (as may be applicable) only come in after consideration of the law of a contracting state. Article 42(1) of the ICISD Convention places overriding consideration on the need to respect and uphold the parties' choice of law and of the law of a contracting state, which in most cases will be municipal law. It also gives overriding consideration to municipal law when it requires that the law of a contracting state, broadly stated, should apply in the absence of choice of law agreement by the parties. It can be argued by this provision that international law only comes in when there is no choice of law and when the law of a contracting state does not apply, or unless its rules of conflict of laws dictate otherwise. These interpretations are right because by virtue of Article 42(2) and (3), a tribunal "may not bring in a finding of *non liquet* on the ground of silence or obscurity of the law" and it does not have the power to decide a dispute *ex aequo et bono* if the parties have not so agreed. Strictly interpreted, under Article 42 of the ICSID Convention, the law of the seat of arbitration does not come in where the parties have made a choice of law.

Thus, a tribunal cannot simply decide to depart from the parties' choice of law because the applicable law will lead to an agreement to arbitrate being invalidated or because equitably and in the sense of judgement of a tribunal, an application of that choice of law will not lead to a party's or the tribunal's expected outcome. It is not the job of a tribunal to validate an agreement to arbitrate against the express dictates of the law chosen by the parties. As the United Nations Conference on Trade and Development (UNCTAD) rightly stated, amiable composition "implies a resort to rules, whereas *ex aequo et bono* involves ignoring them entirely. Either way, neither an amiable composition nor an *ex aequo et bono* clause can be assimilated in any sense to a choice by the parties of non-state rules."[96] Unless a choice of law or rules of conflicts of law of the country chosen dictate otherwise, Article 42 of the ICSID Convention does not retain any discretion by a tribunal to determine the law that is most closely connected to the situation. This is unlike Rome I Regulation, Article 3(3) under which choice of law may prejudice the application of provisions of the law of a country if "all other elements relevant

[96] UNCTAD, *Dispute Settlement: International Commercial Arbitration, Module 5.5 Law Governing the Merits of the Dispute* (New York and Geneva: United Nations 2005) 15.

to the situation at the time of the choice are located in" that country.[97] Article 42 of the ICISD Convention, which gives wide scope for the effect of choice of law, is similar to Article 9 of the Hague Principles on Choice of Law, which reads:

> The law chosen by the parties *shall govern all aspects of the contract* between the parties, *including but not limited to* - a) interpretation; b) rights and obligations arising from the contract; c) performance and the consequences of non-performance, including the assessment of damages; d) the various ways of extinguishing obligations, and prescription and limitation periods; e) validity and the consequences of invalidity of the contract; f) burden of proof and legal presumptions; g) pre-contractual obligations. [98]

Article 8 of the Hague Principles strictly limits the scope of applicable law to the law chosen by parties by excluding rules of private international law from the law chosen by the parties unless the parties expressly provide that rules of private international law shall apply.[99] Similarly, under Article 12 of Rome I Regulation, the law applicable to a contract "shall govern in particular" the interpretation; performance; breach and consequences of breach of obligations; ways of extinguishing obligations, prescription and limitation of actions; and nullity of the contract and the consequences thereof.[100] The use of the phrase "shall govern in particular" suggests that the applicable law may govern other matters other than those specifically mentioned.

In the recent case of *Slowakische Republic (Slovak Republic) v Achmea AV*, decided on 8 March 2018, the European Court of Justice (ECJ) held that an arbitration agreement in an investment treaty between the Slovak Republic and the Kingdom of the Netherlands had an "adverse effect on the autonomy" of European Union law.[101] In the opinion of the ECJ, Articles 267 and 244 of the Treaty on the Functioning of the European Union[102] (TFEU) precluded:

> a provision in an international agreement concluded between Member States [such as the arbitration agreement in the investment

[97] Council Regulation (EC) 593/2008 (n 29), article 3(3).

[98] Hague Conference on Private International Law (n 28) article 9 (emphasis added).

[99] Ibid article 8.

[100] Council Regulation (EC) 593/2008 (n 29), article 12.

[101] Case C-284/16 *Slowakische Republic (Slovak Republic) v Achmea AV* (Judgment of the Grand Chamber, 6 March 2018) [59] <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX-%3A62016CJ0284>.

[102] Consolidated Version of the Treaty on the Functioning of the European Union, 26.10.2012 Official Journal of the European Union C 326/47, <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:12012E/TXT&from=EN>.

> treaty] under which an investor from one of those Member States may, in the event of a dispute concerning investments in the other Member State, bring proceedings against the latter Member State before an arbitral tribunal whose jurisdiction that Member State has undertaken to accept.[103]

The Court reasoned that by concluding the investment treaty, the member parties to it "established a mechanism for settling disputes between an investor and a Member State which could prevent those disputes from being resolved in a manner that ensures the full effectiveness of EU law, even though they might concern the interpretation or application of that law."[104] The ECJ further reasoned that even though the disputes under the investment treaty fell within the jurisdiction of the arbitral tribunal related to the interpretation of the investment treaty and EU law, there was the possibility of those disputes being submitted "to a body which is not part of the judicial system of the EU."[105] The arbitral tribunal established under that investment treaty was thus not situated within the judicial system of the EU and it could not be regarded as a court or tribunal of a Member State within the meaning of Article 267 of TFEU.[106] Only the decisions of courts and tribunals set up by Member States and "situated within the EU judicial system" "are subject to mechanisms capable of ensuring the full effectiveness of the rules of the EU."[107] The instant tribunal was not part of the judicial system of the Netherlands or Slovakia and therefore could not be classified under Article 267 of the TFEU as a court of tribunal of a Member State of the EU.[108]

Article 14 of the Inter-American Convention on the Law Applicable to International Contracts[109] is similar in scope to Article 12 of Rome I Regulation. It has an additional element in Article 14(b) that the "law applicable to the contract" "shall govern principally the rights and obligations of the parties." The obligations of the parties include the obligation to arbitrate. Hence, where the parties have not chosen the law to govern the arbitration clause, the applicable law as to the validity of the arbitration clause would be the Inter-American Convention.

Article 35(1) of UNCITRAL Arbitration Rules is similar in effect to Article 42 of the ICSID Convention. It requires an arbitral tribunal to apply the rules of law designated by the parties for the resolution of a dispute, and it is only when

[103] *Slovak v Achme*a (n 101) [60].
[104] ibid [56].
[105] ibid [58].
[106] ibid [43].
[107] ibid.
[108] ibid [45]–[46].
[109] Inter-American Convention (n 59).

the parties have failed to make the designation that the arbitral tribunal is free to apply the law which it determines to be appropriate. By Article 35(2) of the Rules, an arbitral tribunal may decide as *amiable compositeur* or *ex aequo et bono* "only if the parties have expressly authorized the arbitral tribunal to do so." Thus Article 35(1) and (2) of UNCITRAL Arbitration Rules do not make room for a tribunal to depart from the applicable law chosen by the parties in favour of the law of the seat of arbitration simply because in the tribunal's sense of judgement it is equitable or appropriate to do so. In relation to *BELG v Ghana* analysed above, the governing law to the PPA and the arbitration clause were manifestly closer to the substantive contract and its arbitration clause. Thus, in the *BELG v Ghana* case, which was decided under these Rules, the Tribunal was under obligation to follow the governing law agreed to by the parties, namely Ghanaian law and should not have avoided its application merely because its application would have voided the arbitration clause. If it applied Ghanaian law, it would have found that the arbitration agreement would be invalid. The Supreme Court of Ghana has held in *Attorney-General v Faroe Atlantic Co Ltd,*[110] *Attorney-General v Balkan Energy Ghana Ltd*[111] and *Amidu v Attorney-General Cases*[112] that international business or economic transactions such as power purchase agreements, international loan agreements, and international project implementation agreements that have not been laid before and approved by Parliament as required by Article 181(5) of the Constitution are unconstitutional and void. As such, the State cannot be required to pay damages for breach of such agreements. Such a finding of unconstitutionality would be the result of the parties' choice of law, not of a unilateral application of domestic law by Ghana. It was to avoid these consequences that the Tribunal chose to apply the law of the seat of arbitration in obvious disregard of the parties' freely chosen law.

The position of these arbitration conventions regarding fidelity to the choice of law recognises and gives effect to party autonomy and the need for this autonomy and will of the parties to be respected and upheld.[113] As international arbitration depends on the good faith and domestic institutional support to realise its objectives (for example reliance on domestic courts to enforce arbitral awards), arbitrators should equally respect laws that states have enacted in legitimate exercise of their legislative powers.[114] That obligation to respect and uphold national laws is particularly compelling when the parties to arbitration have chosen a particular national law to govern their relations and when that choice has been made within the limits of mandatory rules. Arbitration being dependent on the agreement of

[110] *Attorney-General v Faroe Atlantic Co Ltd* [2005–2006] SCGLR 271.
[111] *Attorney General v Balkan Energy Ghana Ltd* [2012] 2 SCGLR 998.
[112] *Amidu v Attorney-General* [2013–2014] 1 SCGLR 112; [2013–2014] 1 SCGLR 167.
[113] Hochstrasser (n 18) 58 and 84.
[114] ibid 85.

the parties to submit their disputes for settlement by arbitration, arbitrators derive their authority from such agreement and the exercise of their powers must equally be limited by the express and implied effect of the arbitration agreement and choice of law made by the parties. Arbitrators must not substitute their own view as to what the parties should have agreed to for what the parties have actually or impliedly agreed to. Arbitrators should decide within the scope of the agreement to arbitrate, the chosen law to be selected, and should not wish for the parties what they have not wished for themselves.[115] Professor Buys rightly argues this point when she states that "[i]n light of the contractual nature of arbitration, parties to a commercial arbitration agreement ought to be assured that their choice of the applicable law will be respected as well."[116] In other words, according to Professor Buys:

> [as] arbitrators are appointed by the private agreement of the parties… their allegiance should be to seeing that the terms of that private agreement are carried out. The arbitrator's duty to respect the wishes of the parties as expressed in their written agreement extends to respect for the parties' choice of law in a commercial transaction.[117]

Arbitral respect for the parties' choices as reflected in the agreement to arbitrate is not only consistent with the underlying principles of freedom of contract and party autonomy, it could also lead to reviewing courts respecting party autonomy and upholding decisions of arbitral tribunals, subject only to overriding public policy considerations and mandatory national rules.[118]

Despite the need to keep faith with party autonomy and the parties' choice of law made in exercise of that autonomy, there are cases, as reflected in *BELG v Ghana*, when one party becomes dissatisfied with the choice of law that the parties have agreed to and may seek to have the arbitrator apply the law of a different country such as that of the seat of arbitration. The party objecting to the application of the choice of law might argue that there does not exist a reasonable or substantial connection between the law chosen and the parties' transaction. The party may also be seeking simply to avoid the natural consequences of an unfavourable outcome against it that may result from the application of the chosen law. For example, in *BELG v Ghana*, BELG requested the Tribunal to "refuse to apply the choice-of-law clause to the arbitration clause because Ghanaian law

[115] Buys (n 26) 67–68.
[116] ibid 69.
[117] ibid.
[118] ibid 71.

discriminates between national and international transactions to which Ghana is a party and thwarts the Parties' true intentions to arbitrate."[119] It is further argued that if the Tribunal proceeded to apply conflict of laws rules, "the conflict of law rules of the seat of arbitration should be applied in determining the substantive law applicable to an arbitration agreement."[120]

In most cases, the tribunal will turn to the law of the seat of arbitration because it is traditionally taken that the designation of a place as the seat of arbitration is treated as consent to the procedural law of that place and as evidence that the substantive law of the seat applies.[121] In *BELG v Ghana*, the Tribunal held that the parties' agreement to dispute settlement before the Permanent Court of Arbitration "is an indicator that the Parties intended to remove questions relating to dispute resolution – as opposed to the substantive performance of the contract – from the place of either Party, to a neutral forum."[122] This position does not reflect the agreement of the parties. The Parties had expressly agreed to UNCITRAL Arbitration Rules for the conduct of arbitration. Even if the governing law to the main contract did not apply to the arbitration agreement, UNCITRAL Arbitration Rules did and ought to have been the reference for deciding whether the arbitration agreement was valid.

The parties to a choice of law agreement should not be allowed to evade their contractual obligation to observe that agreement if they are challenging the applicability of the chosen law in disguise—that is, as a camouflage to avoid the consequences attached to the application of the law they have freely agreed. If the chosen law cannot be applied for some substantive and legitimate reason, a tribunal should apply the conflict of law rules that accompany the substantive law chosen by the parties. This is because in the absence of a specific definition of the scope of the chosen law by the parties themselves, the chosen law must be taken to include its conflict of law rules.

The conventional theory that the choice of a place as the seat of arbitration gives reason to disregard the chosen law in favour of the law of the place of arbitration needs to be revisited. The place of arbitration is simply chosen as a place of arbitration for purposes of dispute resolution. If the parties really consider the law of the place of arbitration as neutral, they probably will go ahead to expressly make a choice of that law instead of choosing the law of one of the parties as the governing law. Therefore, in the absence of the parties expressly giving some role for the law of the seat of arbitration in the parties' dispute resolution, the

[119] *BELG v Ghana* (n 37) [130].
[120] ibid [131].
[121] Buys (n 26) 71.
[122] *BELG v Ghana* (n 37) [142].

law of that place is in no different position than the law of any other place which could equally have been chosen as the seat of arbitration or which has some other connecting fact but has not been chosen. The seat of arbitration has interest in the way justice is administered within its territory. That country's legal system, however, may also recognise party autonomy and its laws must not be applied to defeat party autonomy which has been legitimately exercised.

It may be argued that there may be no connection between the parties to the matrix contract and the governing law they have chosen.[123] While there are other important factors that can connect the parties and their transaction, it must to be borne in mind that the very purpose of choice of law is to connect the chosen law, the parties, their transaction, and their disputes. Thus, choice of law is a connecting factor. Indeed, this is the primary function of making choice of law. The parties would have made that choice of law well aware of other connecting factors. In choosing the laws of a particular jurisdiction, the parties mean to exclude all other factors that could plausibly be connected to their transaction and dispute. Therefore, it seems there can hardly be a lack of connection between a transaction or dispute and the chosen law. If the parties have the freedom to contract as well as the autonomy to choose the law to govern their contract and they have legitimately done so, there must be compelling and fundamental factors to ignore that choice as a connecting factor. Such a compelling factor cannot be that the chosen law will invalidate the underlying agreement or agreement to arbitrate. Such a compelling factor does not also include a tribunal's wish or desire to do justice beyond the terms of laws governing the exercise of its powers—that is, if the tribunal is not authorised by the parties or the applicable arbitration convention to exercise its powers as *amiable compositeur* or *ex aequo et bono*.

V. Conclusion

Professors Elliot Cheatham and Willis Reese rightly argue that: [124]

> All laws, be they statutory or judge-made, are the products of policies which, in turn, are deemed of sufficient importance to warrant their embodiment into law. A choice of law decision is, therefore, of real concern to the states involved, since in net effect it determines whose policy shall prevail in the particular case. This consideration dictates that the law of the state with the dominant interest, should normally at least, be applied.

[123] Born (n 20) and Buys (n 26).
[124] Cheatham and Reese (n 2).

I argue that if the interest of the state is a relevant criterion, "the state with the dominant interest" to have its law applied in a dispute involving parties *who have made a contractual choice of law* is the state the laws of which have been chosen as the applicable law to the contract. International arbitration tribunals should faithfully apply the law chosen by the parties because the choice of applicable law expresses the will and intention of the parties that that law, and not any other, shall apply to their transaction and dispute. In furtherance of the object of predictability and certainty as to the applicable law to the terms of the underlying contract and the rights and obligations of the parties deriving from that contract, the chosen law has express and implied application to both the matrix contract and arbitration clause embedded in it unless otherwise expressly indicated by the parties.

Therefore, if that choice of applicable law has been legitimately made, then effect must be given to the parties' will and intention. If the parties have made a choice of law, a tribunal and a party to the choice of law agreement must not seek overtly or covertly to avoid the application of the chosen simply because of unpalatable, undesirable or negative outcomes that will naturally arise from the application of the chosen law. Accordingly, I adopt Professor Park's perspective that:

> No one opts for an unfair result applied to himself. However, it is rarely possible to predict in advance of the dispute who will get the rough side of the law, since the contours of the controversy do not exist. For this reason, parties to commercial transactions agree to "play by the rules," aware that application of the rules will not always produce agreeable results. It is not irrational to assume that businessmen desire the application of rules of law as an accepted calculus of justice, even though those rules lead to consequences that could be described as unfair. [125]

[125] Park (n 16) 662.

The Destination-Based Corporation Tax: A Solution to Formalism in Source-Based Taxation?

*Shaun Matos**

Introduction

The taxation of multinationals has become one of the biggest issues in modern discourse. With fierce market competition, it is unsurprising that businesses have started to view tax obligations as a cost to be minimized rather than as a social or political obligation. Such behaviour is problematic not only for the effects it has on government revenues, but also for its link to other problems such as competition and distortions. Many of these problems relate to the mechanics of source-based taxation and are linked to the fact that the definition of 'source' is unclear, slippery, and open to manipulation. The source-based system as a whole is very formalistic, in that it looks only to the place where multinationals have chosen to locate their profits rather than to the place where profits are actually generated. This formalism will be explored further in this article.

Yet, if we believe certain commentators, we may not have to give up hope on finding a tax system which provides a solution to these problems. A number of academics have built on the work of the Meade Committee to argue in favour of a Destination-Based Cash Flow Tax (DBCFT), a version of which was once supported by the Republican Party (GOP). It is argued that, if universally adopted, this approach would alleviate many of the problems that the source-based system currently produces. Much of this rests on 'destination' being a more certain and less formalistic concept than 'source'. This paper will examine

* LLM Candidate at University College London, BA (Cantab). With thanks to Dr Dominic de Cogan for his initial supervision of the article, Professor Rita de la Feria for her additional comments and the editors for their thoughts and advice.

this claim, drawing heavily on VAT literature. It will be argued that, whilst the destination principle does indeed deal with some of the major problems of the source-based system, there are still a number of practical problems inherent in any destination-based system. Although there will be a brief mention of the economic debates and some of the wider legal issues, the focus will be on whether the tax and the destination principle it relies upon, can be implemented in a less formalistic manner. The potential issues can be seen most clearly when considering multi-location entities and branch mismatch arrangements. The first part will briefly explore the problems of source-based taxation and the solutions which the DBCFT purports to provide. The second part will then consider whether this system could actually be implemented.

I. Part I

A. The basics: source and destination

Before proceeding any further, it is necessary to develop an understanding of the difference between 'source' and 'destination' as tax bases. Griffith, Hines, and Sørensen define source-based taxation as the system in which "capital is taxed only in the country where it is invested"[1] resulting in a tax on investment. Hence, if Company X invests in Country A and sells into Country B, a source-based system requires that X only pays tax in A. The authors give a brief explanation of the economic effect of such taxation in a small open market. As they argue,

> if the domestic government imposes a source-based business income tax, the pre-tax return to domestic investment will have to rise by a corresponding amount to generate the after-tax return required by international investors. Hence, domestic investment will fall and capital will flow out of the country until the pre-tax return has risen sufficiently to compensate investors fully for the imposition of the source tax. Thus, the incidence of a source-based capital tax falls entirely on the immobile domestic factors of production (land and labour).[2]

It should be noted that the source-based system is not the only model currently in use. There is also a 'residence-based system' under which companies are taxed by their country of residence on their worldwide profits, and most

[1] Rachel Griffith, James Hines, and Peter Birch Sørensen, "International Capital Taxation". in James A Mirrlees and Stuart Adam (eds), Review, *Dimensions of Tax Design: The Mirrlees Review* (Oxford University Press 2010) 925.

[2] ibid.

national tax codes adopt a combination of both systems to maximize their tax receipts.[3] The source-based model, however, is predominant[4] and is the one that the proponents of the DBCFT focus on. The residence-based system will therefore not be discussed any further.

There are three core elements to the DBCFT. The basic idea behind the 'destination' element is that, similar to a VAT, tax is levied based on the location of the customer rather than that of the supplier.[5] There is, however, one crucial difference between the DBCFT and the VAT. Even though VAT is usually conceived as being a tax on consumption, in the majority of cases it will be impossible to know where the good or intangible is consumed (as opposed to purchased).[6] Hence, when VAT is charged on a transaction, the location of the customer is merely a proxy for the place of consumption. Often there will be numerous intermediaries in a chain of transactions before the product is finally consumed, meaning that identifying the location of a customer may not be enough. Other proxies, such as the location of the goods, are therefore used in conjunction.[7] The destination of the customer is, however, still critical to VAT's staged collection mechanism.[8] The DBCFT, on the other hand, is not a tax on consumption, but on either profit or imports. Destination here is therefore not a proxy for consumption,[9] in the sense that in most cases there is no need to use a combination of proxies to determine the appropriate place of taxation. This will be discussed in Part II section B.

The second part is the cash flow element. This is based on a proposal initially put forward by the Meade Committee, and can be explained as follows:

> [A] cash flow tax applies to net receipts arising in the business. Receipts are included in the tax base when payment is received and expenses are recognized when payment is paid. The tax base in any given period is the former less the latter. The most significant difference in the timing of the inclusion of receipts and expenses in the base, compared with most existing corporate tax systems, is

3 See, for example, Section 5(1) of the Corporation Tax Act.

4 See Article 7 of the Organisation for Economic Co-operation and Development (OECD), *Model Tax Convention on Income and on Capital: Condensed Version 2017* (OECD Publishing 2017).

5 Alan Auerbach, Michael Devereux, and Helen Simpson, "Taxing Corporate Income". in James A Mirrlees and Stuard Adam (eds), *Dimensions of Tax Design: The Mirrlees Review* (Oxford University Press 2010) 839 and 883. See also Section 9.6.3 more generally.

6 Organisation for Economic Co-operation and Development (OECD), *International VAT/GST Guidelines* (OECD Publishing 2017) 17.

7 Michael Devereux and Rita de la Feria, "Designing and implementing a destination-based corporate tax" (2014) *Oxford University Centre for Business Taxation* Working Paper Series 14/07, 16 <http://eureka.sbs.ox.ac.uk/5081/1/WP1407.pdf> accessed 1 May 2018; Rebecca Millar, "Cross Border Services - A Survey of the Issues". in Richard Krever and David White (eds), *GST in Retrospect and Prospect* (Thomas Bookers Ltd 2007) 322-323.

8 OECD, (n 6) 38.

9 Devereux and de la Feria, (n7) 10.

> that under cash flow taxation even capital assets that are typically depreciated over time are immediately expensed.[10]

There are two variations: an R-based system which ignores transactions involving financial assets and liabilities, and an R+F-based system which includes financial as well as 'real' flows.11 There is a significant amount of academic discussion on this part of the tax.12 Because of space constraints and the focus of other literature, however, this paper will only focus on the 'destination' aspect of the tax.

This leads to the final element of the tax which is key to explaining how the previous two aspects operate in practice. Under the border-adjustment mechanism, the liabilities of a taxable entity differ depending on whether the entity is an importer or an exporter. If it is an exporter, the mechanism operates to reduce the business' tax burden by zero rating exports, with the result being that they are not included in the business' taxable profits. If, on the other hand, the business is an importer, the mechanism operates so as to impose a tax on the imported goods. This could be done in one of two ways: (a) by directly taxing the import and allowing the importer to deduct the price of the import from its tax base, or (b) by not taxing the import but not allowing a deduction for the cost of the import.[13] In either way, the effective tax base is domestic sales minus domestic costs.[14]

The potential benefits that many find in the tax when compared with a source-based system will be explored below, but it is, at this point, worth briefly exploring the potential economic effects of the tax. Those in favour argue that the DBCFT would be trade-neutral for any individual state because, even though exports would become cheaper and imports more expensive, an increase in the price of labour and an appreciation of the exchange rate would result in

[10] Alan Auerbach, Michael Devereux, Michael Keen, and John Vella, "Destination-Based Cash Flow Tax" (2017) Oxford Legal Studies Research Paper 14/2017, 9 <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2908158&download=yes> accessed 1 May 2018.

[11] ibid.

[12] David A Weisbach, "A Guide to the GOP Tax Plan - The Way to a Better Way" (2017) University of Chicago Coase-Sandor Institute for Law & Economics Research Paper No. 788, 288 <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2893224&download=yes> accessed 1 May 2018; James E Meade, *The Structure and Reform of Direct Taxation* (Allen and Unwin 1978).

[13] Auerbach, Devereux, Keen, and Vella, (n 10) 14; Weisbach, (n 12) 27-28. Despite the fact that there are two ways in which the tax could operate, much of the analysis in the papers referred to seems to assume that, in relation to businesses at least, the second version of the DBCFT would be adopted. Therefore, unless specified otherwise, the discussion below assumes that it is the second model of the DBCFT that is being adopted.

[14] The Economist, "How America's border-adjusted corporate tax would work" (*The Economist* 13 February 2017) <https://www.economist.com/blogs/economist-explains/2017/02/economist-explains-9?fsrc=scn/tw/te/bl/ed/> accessed 1 May 2018.

an equilibrium.[15] It is difficult to predict with certainty whether prices and the exchange rate would adjust in this way, especially in the case of unilateral adoption, but it is sufficient to say that it is, at least theoretically, possible for the tax to have no effect on the balance of trade.

Before moving on, it is worth highlighting the distinguishing features of the DBCFT. Firstly, exports are not included in the tax base (and it is possible to get a tax credit).[16] Secondly, businesses are unable to deduct the cost of imports from the tax base unless a direct tax is imposed on the import. Thirdly, businesses are able to deduct labour costs from the tax base. It is these features, and especially the first two, which are relied upon to eliminate the formalism in the source-based system.

B. The problems of a source-based corporation tax

The problems of a source-based corporation tax are numerous, so this section will focus on those in which the impact of formalism is the clearest.

1. Avoidance

As mentioned in Part I section A, the opportunity for avoidance is perhaps the gravest issue that the current system faces, especially considering the fact that a number of other problems flow from it. The literature focuses on three main methods of avoidance, so it is these methods that will be considered here.[17]

The first method, namely debt shifting, has more to do with the cash flow aspect of the DBCFT (see Part I, section C(1)) and it will be considered only briefly. The basic idea is that a company or group with establishments in multiple jurisdictions can strategically lend money between jurisdictions to take advantage of the most favourable rules regarding tax relief and interest income. The result is that multinationals will borrow in high rate jurisdictions to reduce the taxable profits there.[18] Such groups may also then use subsidiaries in low rate jurisdictions to lend to affiliates in high rate jurisdictions as the deductible interest payment in the high rate country will likely exceed the taxable interest receipt in the low rate jurisdiction. It is unnecessary to undertake an in-depth analysis of this example to see the formalism inherent in the system; multinationals can easily use debt

[15] Auerbach, Devereux, Keen, and Vella, (n 10) 17-21.

[16] The presence of tax credits may cause further problems in itself. For this, see Reuven S Avi-Yonah and Kimberly A Clausing, "Problems with Destination-Based Corporate Taxes and the Ryan Blueprint" (2017). University of Michigan Law & Economics Research Paper 16-029, 17-18 <https://ssrn.com/abstract=2884903> accessed 1 May 2018; Weisbach, (n 12) 55.

[17] Auerbach, Devereux, Keen, and Vella, (n 10) 27-30.

[18] ibid.

to locate their profits in the most tax-efficient way because of the inability of the source-based system to look beyond the locations chosen by multinationals.

Manipulation of intra-group pricing is the second method outlined in the literature and is much more significant for the purposes of this paper. It occurs when companies, which are related to or part of the same multinational group, trade with each other and misstate the price of the relevant goods. For example, if State A has a higher tax rate than State B, there is an incentive to understate the true price of goods transferred from a company in A to a company in B because B's tax relief will exceed the tax levied on A's sale.[19]

The consequence is that the company in the high tax jurisdiction seems to be significantly less profitable. It is easy to see that this method is again largely inefficient because of the formalism and vagueness of 'source' as a basis for taxation. Despite the fact that tax is supposed to be levied where the profit-generating activity is carried out, the formal and artificial location of the profits is the focus of a source-based inquiry. Multinationals therefore are free to enjoy the best of both worlds, meaning that they are able to locate their production centres in the jurisdictions which have the best resources while paying tax in whichever jurisdiction is most beneficial. Hence, a more substantive approach is needed.

The final method pointed out by supporters of the DBCFT is the placement of valuable intangible assets in low tax jurisdictions. The parts of the multinational located in high tax jurisdictions pay royalties in return for the use of assets, giving tax relief at the high rate and giving rise to a tax receipt in the low tax jurisdiction.[20] Again, it is not difficult to see that the uncertainty inherent in the concept of 'source' is responsible for this. Whilst, on a formal level, the asset appears to be generating profit in the low tax jurisdiction, it is often the case that the asset will have been developed elsewhere and then sold to a related company in a low rate jurisdiction. This means that the true source of the revenue, namely the work done to produce the asset, will be hidden.

Before moving on, it is worth noting that a number of measures have been taken to address these issues. One measure is the arms-length pricing rule, which stipulates that related parties must deal with each other as if they were entirely independent to prevent manipulation of intra-group prices. At first glance, this does appear to prevent parties from abusing the formalism of the 'source' concept, but there are a number of problems with this solution in practice. The Mirrlees Review points out that this principle can be difficult to apply where transactions involve highly specialised products which may not be traded by other commercial parties. Moreover, in extreme cases, the principle may break down completely

[19] ibid.
[20] ibid.

where the product is unique and there are no comparable transactions taking place in the normal market.[21] In such cases, it is difficult to identify the true market price of the product and hence, the places where profits are actually being generated. Indeed, Mirrlees goes even further to suggest that the principle may be generally flawed because it cannot take into account other advantages multinationals will likely enjoy from trading with related parties.[22] Hence, it is clear that, despite efforts to close these avoidance channels, the formalism of a source-based tax means that companies are able to conceal the true location of their profits with relative ease.

2. Efficiency

Connected to these avoidance mechanisms is the distortionary effect source-based taxation can have. Auerbach, Devereux, and Simpson set out the stages of an investment decision and point out how each stage can be affected by tax considerations.[23] For example, different rates of corporation tax can affect a company's decision whether to expand by producing more domestically or abroad. Further, as mentioned, source-based taxation can also be an important factor in deciding where profits should be located and how a multinational should be structured. It is true that tax is not the only factor which affects business decisions but, given the significant empirical evidence in favour of this analysis,[24] one is led to the conclusion that the formalism of a source-based system can, and does, result in significant distortions.

3. Tax Competition

A further issue concerns the competition generated between states. As authorities become aware of avoidance mechanisms and the distortionary effect of taxation, it is only natural that they try to maximise their revenues by creating favourable environments. There is, however, a real risk of a race to the bottom. Since 1978 the main rate of corporation tax in the UK has fallen from 52% to 19% and is scheduled to drop further. Similar falls can also be seen in other countries.[25] Of course, changes in tax rates cannot be viewed in isolation; one must

[21] James A Mirrlees and Stuart Adam, *Tax by design: the Mirrlees Review* (Oxford University Press 2011) 434-436.

[22] ibid.

[23] Auerbach, Devereux, and Simpson, (n 5) 853-855.

[24] ibid.

[25] Stuart Adam, James Browne, and Christopher Heady, "Taxation in the UK". in James A Mirrlees and Stuart Adam (eds), *Dimensions of Tax Design: The Mirrlees Review* (Oxford University Press 2010) 25; UK Government, "Rates and allowances: Corporation Tax" (UK Government, 1 April 2018) <https://www.gov.uk/government/publications/rates-and-allowances-corporation-tax/rates-and-allowances-corporation-tax> accessed 1 May 2018.

remember that the fall in rates has been accompanied by a broadening of the tax base and that the share of corporation tax receipts in total revenue and GDP has not significantly changed.[26] Further, decisions to reduce the rate of corporation tax may also have been influenced by a desire to remove capital controls to welcome international investors. Nevertheless, efforts to compete on rates and more widely, with other states for tax revenues may lead to the creation of further distortions and the usage of avoidance mechanisms. It is easy to see how a vicious circle can be formed over time. More broadly, in a globalised world where cooperation is key to tackling issues such as avoidance, a system which incentivises competition is evidently not a good thing.

A. The promises of the DBCFT

While the Organization's for Economic Co-operation and Development (OECD) Base Erosion and Profit Shifting (BEPS) project attempts to provide a solution to the problems outlined in section B, those in favour of the DBCFT suggest that their model can provide a neater and simpler solution. The model does appear to be successful in avoiding the problems outlined because a destination-based tax system avoids some of the formalism and uncertainty inherent in the source-based system. Given the close link between avoidance and other issues outlined above, this section will mainly focus on the DBCFT's solutions to avoidance mechanisms.

1. Avoidance

As stated, preventing the shifting of debt is more closely related to the cash-flow element of the tax. In short, the cash-flow element can come in a number of guises, the most likely of which to be adopted are the R-based system and the R+F-based system. Under the R-based system, there would be no consequences for either interest payments or interest receipts.[27] Hence, this method of avoidance would simply not exist. The R+F base, on the other hand, applies to "all net financial inflows other than equity transactions with [...] shareholders".[28] The exact mechanics of the R+F version are complex and not relevant to the main argument of this paper.[29] For present purposes, it is enough to say that the debt-shifting channel would not be available under the R+F system either.[30] The cash-

[26] Adam, Browne, and Heady, (n 25) 8.
[27] Meade (n 12); Auerbach, Devereux, Keen, and Vella, (n 10) 9-10.
[28] Auerbach, Devereux, Keen, and Vella, (n 10) 45.
[29] ibid, 45-63.
[30] ibid, 27.

flow element of the tax, therefore, appears to eliminate the formalism of the source-based system.

Of more relevance is the manipulation of intra-group prices. The destination principle means that an exporting company faces no domestic taxes as exports are zero-rated and the importing company will be liable instead. Under the first version of the DBCFT, this liability will take the form of an import tax accompanied by a deduction of the price of the import from the company's tax base. We are told that these "effects exactly cancel out, making the value of the import irrelevant for tax purposes".[31] Under the second version of the DBCFT, where imports from taxable businesses are simply excluded from tax calculations, the transaction between the related parties is free of tax.[32] Cross-border transactions, therefore, apparently have no tax implications for either party in the sense that a multinational cannot manipulate its tax liability. The authors also point out this 'netting out' of business-to-business (B2B) transactions removes the opportunity under the less popular formulary apportionment system for a highly profitable company to sell its products in an arms-length transaction to a less profitable company in a low-tax jurisdiction which would then re-sell the goods to consumers in a high tax jurisdiction and pay the high rate of tax on its low profits.[33] The foregoing discussion appears to show that the formalism, which plagues the source-based system, would not be an issue if it is possible to implement the DBCFT: companies simply wouldn't be able to misrepresent the location of their profits in the same way.

The final method mentioned above concerning the location of intangible assets would also be eliminated. Regardless of the source of the asset, it will be taxed at the full rate in the jurisdiction in which it is sold, with the tax in the destination jurisdiction again being deductible for the receiving company. As the tax liability is the same wherever the asset is located and it is not possible to mask the true destination of the asset, the DBCFT seems to rule out this channel of avoidance as well.[34]

2. Efficiency and Competition

As noted in section B, the source-based system has considerable distortionary effects, but this does not seem to be an issue under the DBCFT. One reason for this relates to the preceding discussion as there is no incentive for businesses to shift their profits. The second reason is that consumers are relatively immobile and thus they are unlikely to change locations because of tax rates. It follows that there is no

[31] ibid, 29.
[32] ibid, 28.
[33] ibid.
[34] ibid, 29-30.

incentive for states to compete on tax rates.[35] Auerbach, Devereux, Keen, and Vella note that one potential issue is that firms may have an incentive to locate expenses in a country with a higher rate of DBCFT because that gives them the option to deduct expenses from profits. After carrying out an in-depth economic analysis of the consequences of doing this, however, the authors suggest that changes in demand from both the lower and higher rate jurisdictions would lead to an upward pressure on the currency of the higher rate jurisdiction. The value of profits earned, therefore, increases and sets off the higher rate of tax. If such analysis is correct, it shows that, once the formalism and uncertainty of the source-based system are removed, many of the problematic consequences that flow from it seem to disappear. Instead, business decisions will be driven by other economic factors and tax policy will be implemented in an international environment characterised by cooperation rather than competition.

II. Part II

A. Some basic problems

Given the analysis in Part I section C, it is fair to say that the DBCFT does indeed resolve some of the problems which currently exist in the source-based system. Nevertheless, it has some issues of its own. Before moving onto a more detailed analysis of the destination principle, there are some initial problems with the DBCFT that should be outlined in the interests of completeness. Indeed, many of these problems may be in and of themselves big enough to make the implementation of the tax difficult, if not impossible.

1. WTO Rules

The first major problem is that the tax may be viewed as incompatible with the World Trade Organization's (WTO) rules. The WTO Subsidies and Countervailing Measures Agreement and the Article 16 of the GATT outlaw export subsidies, the definition of which appears to cover a border-adjustable direct tax, stipulating that: "(e) The full or partial exemption, remission, or deferral specifically related to exports, of direct taxes (58) or social welfare charges paid or payable by industrial or commercial enterprises (59) (Annex 1)".

Footnote 58 then defines 'direct tax' as "taxes on wages, profits, interests, rents, royalties, and all other forms of income, and taxes on the ownership of real property". This seems to straightforwardly cover the DBCFT. It is true that

[35] Ibid, 23-26.

VAT, as an indirect tax, is not incompatible with this rule. Nonetheless, the major difference between the two is that the DBCFT allows labour costs to be deducted, whilst VAT does not. This may seem like a minor difference but Avi-Yonah and Clausing show that it results in exports being subsidized by the state in a way they would not be under VAT.[36] There are three further factors that point away from the DBCFT being treated as analogous to VAT: (a) the overall tax base being the profit of the enterprise rather than the value added; (b) the intention that the tax being borne by the owners of capital rather than final consumers; and (c) the fact that the tax is levied based on where the taxpayer delivers the service or goods rather than where consumption occurs.[37] It may be possible to get around this problem by taking the deduction for labour costs outside of the tax. How this could be done and what effects it could have on the rest of the tax system is, however, unrelated to the destination principle and so, outside the scope of this paper.[38]

2. Currency Exchange

As we have seen in Part I section C, the ability of the DBCFT to eliminate the distortionary effects tax can have on investment decisions seems to be one of its major advantages, even where tax rates are not the same across jurisdictions. This depends, however, on upwards currency pressure cancelling out the effect of the higher rate of tax[39] which is a proposition not all economists accept. Gale, for example, points out that such analysis assumes that capital market flows would not be affected, but argues that this is in fact very likely to happen.[40] He also points out that there are a number of other factors that make it difficult to estimate the effects of border adjustment on the exchange rate in practice; these include slow adjustment of prices and pre-existing contracts affecting exchange dynamics.[41] Rogoff also criticises this kind of modelling, suggesting that it is extremely difficult to use structural modelling to predict exchange rate volatility.[42]

Avi-Yonah and Clausing point to a number of other issues, such as the fact that many countries fix their exchange rates making it difficult to predict the effect

[36] Avi-Yonah and Clausing, (n 16) 5-14.

[37] ibid.

[38] There is also a range of other WTO issues, such as discrimination of like products which is forbidden by GATT, Article II, [2] and [4].

[39] Auerbach, Devereux, Keen, and Vella, (n 10) 25-26.

[40] William G Gale, "Understanding the Republicans' corporate tax reform" (*Brookings* 10 January 2017) <https://www.brookings.edu/opinions/understanding-the-republicans-corporate-tax-reform/> accessed 1 May 2018.

[41] ibid.

[42] Kenneth Rogoff and Martin Feldstein, "Perspectives on Exchange Rate Volatility". in Martin Feldstein (ed), *International Capital Flows* (University of Chicago Press 1999) 444.

on the it.[43] If such doubts are correct, this significantly undermines the efficacy of the DBCFT. Not only does it impair the neutrality of the tax, but it can also have an impact on trade and consumer prices. Weisbach argues that, even if the models are correct and the adjustment would take place, the transition could be slow and rocky, potentially still having a significant negative effect on importers.[44] The uncertainty surrounding this aspect of the proposed tax makes it difficult to envisage it being implemented by any but the bravest of governments.[45]

A. Learning from VAT

There are a number of differences between the DBCFT and the VAT. The VAT is, in theory, an indirect tax on consumption whereas the DBCFT is a direct tax, either on corporate profits or imports. Consequently, the DBCFT gives relief for labour costs whereas the VAT does not because the value added is equivalent to the profit earned plus the amount paid for labour. Moreover, the DBCFT is not reliant on proxies in the same way as the VAT is, and the collection mechanisms for each differ. Most VATs rely on the invoice credit method under which the supplier charges the customer VAT on each transaction.[46] If the customer is a VAT-registered business, it can credit this VAT against output taxes charged on sales. The importance of this collection method is difficult to overstate, something that can be seen by the fact that the OECD report calls it "the central design feature of a VAT".[47] In contrast, the DBCFT uses the subtraction-based method which looks much more like a traditional corporation tax.[48] Under the subtraction-based method, an annual account is taken of sales and input costs (labour costs are not included where the tax is VAT but are where it is the DBCFT) are deducted.[49] This distinction will become important in the following discussion. It is important to emphasize, at this point, the sense in which destination will be used. As both a VAT and the DBCFT tax imports rather than exports, we are concerned mainly with the jurisdiction of the imported product rather than the jurisdiction of any later exports. Despite the differences, the destination of the goods or service is vital to both. We can, therefore, learn a lot from the mechanisms in which the concept has operated in the VAT context and the problems that have arisen there. Much of the rest of this paper will explore the inconsistent implementation of the destination

[43] Avi-Yonah and Clausing, (n 16) 11.

[44] Weisbach, (n 12) 46.

[45] These examples are illustrative only. There is also a range of further issues such as the effect on tax treaties. See Weisbach (n 12) and Avi-Yonah and Clausing (n 16) for further discussion.

[46] OECD, (n 6) 15.

[47] ibid, 14

[48] Auerbach, Devereux, Keen, and Vella, (n 10) 16; Meade, (n 12) 86-88.

[49] Auerbach, Devereux, Keen, and Vella, (n 10) 16.

principle across jurisdictions. It will be argued that, despite the apparent success of the DBCFT in closing avoidance mechanisms which exist under the source-based system, it is unable to implement the destination principle in a way that does not create new avoidance mechanisms. This can be seen most clearly in the context of transactions involving services and intangibles. Whilst solutions may be available in the VAT context, they cannot be readily applied to the DBCFT. Discussion of this problem will show that in practice the concept of 'destination' is not as certain and substantive as the preceding section would suggest and may cause significant problems if the DBCFT were to be implemented.

1. Defining and Implementing Destination

A broad distinction can be drawn between two types of trade: trade of goods, on the one hand, and the trade of services and intangibles, on the other. The treatment of the destination principle with regards to goods can be dealt with fairly briefly because the existence of border controls and fiscal frontiers means that it is simple to identify the place of delivery.[50]

Trying to find a way to consistently implement the destination principle with regards to services and intangibles is far more difficult because, unlike goods, the nature of services and intangibles means that they cannot be subject to border controls.[51] As the traditional approach of determining the place of consumption by the location of the provider is clearly insufficient in a globalised world, a new way must be found. Keen and Hellerstein helpfully divide this category up into 'tangible services' and 'intangible services'.[52] The former category refers to services "with respect to which sensible proxies for consumption can be identified by reference to the location of the performance of the services, or their link to the location of the property, rather than the location of the customer". While determining which jurisdiction collects VAT in such cases can be relatively straightforward where the customer's location is used as a proxy; Millar points out that this is not always the case. Sometimes other proxies, such as the location of relevant goods, the location of immovable property, and the place of effective use and enjoyment are used instead.[53] Given that "VAT regimes differ in the way that they combine the proxies, the order of their application and the priority given to each proxy",[54] there appears to be a real risk of double or nil-taxation. Despite these problems, this category

[50] OECD, (n 6) 17.
[51] ibid.
[52] Walter Hellerstein, and Michael Keen, "Interjurisdictional Issues in the Design of a VAT" [2010] 63(2) Tax Law Review 372.
[53] Devereux and de la Feria, (n 7) 322-323.
[54] Hellerstein, and Keen, (n 52) 374.

of services is unlikely to be an issue for the DBCFT because, unlike the VAT, the DBCTF is not concerned with the location of consumption. The location of the customer is *always* the focus of the DBCFT so there is no need to become concerned with other proxies.

The second category concerns services which, in contrast, do not have a connection with any identifiable physical location. Unlike 'tangible services', there is no location or physical element which can be relied on to provide a means of identifying the place of consumption in place of border controls. There are two distinct scenarios here: business-to-business transactions and business-to-consumer transactions.[55] Within business-to-business transactions, there is a further divide between sales to single location entities (SLEs) on the one hand, and sales to multiple location entities (MLEs), on the other. Given that in business to consumer contexts it is relatively straightforward to identify the location of the consumer, this issue will not be considered further.[56]

Moreover, where the customer in a business to business transaction is a SLE, the solution is straightforward.[57] The problem, however, is more complex where the customer is an MLE. In such circumstances, it is often unclear which establishments have used the service or intangible and hence, which jurisdictions have taxing rights.[58] This means that, unlike in the discussion above, the problem is not unique to VAT. It has nothing to do with destination being a proxy for consumption or the use of different combinations of proxies. Rather, it is about finding the real destination of services or intangibles in any given transaction, meaning that it is inherent in *any* destination-based system. Therefore, unless there is a single set of rules that determines the destination of services and intangibles, there is a risk that the universal adoption of the DBCFT may lead to gaps and overlaps in the taxation of corporations. This problem is acknowledged by the supporters of the DBCFT but it is dealt with fairly briefly by simply recommending the adoption of the OECD's guidelines.[59] It will become clear that this lack of discussion, in particular on the issue of whether the OECD's guidelines are applicable in a DBCFT context, means that the tax's formalism and susceptibility to avoidance have not been sufficiently recognized.

The OECD's 2017 International VAT/GST Guidelines attempt to deal with the issue of defining and implementing 'destination' in a VAT context. Guideline 3.2 says: "[f]or the application of Guideline 3.1 [which recommends the use of the destination principle] for business-to-business supplies, the jurisdiction in which

[55] OECD, (n 6) 38.
[56] ibid, 64-72 for further detail.
[57] ibid, 44.
[58] ibid, 45.
[59] Devereux and de la Feria, (n 7) 17; Auerbach, Devereux, Keen, and Vella, (n 10) 79.

the customer is located has the taxing rights over internationally traded services or intangibles". Guideline 3.3 supplements this by stating that, "[...] the identity of the customer is normally determined by reference to the business agreement". In situations where the customer is a SLE, this rule is easy to implement: all the supplier needs to do to ensure the export is free of tax is to identify and demonstrate who their customer is and where they are located. This would be easy for the parties to do and for tax authorities to monitor.

The picture is, however, more complicated where the customer is an MLE, as it is necessary "to determine which of the jurisdictions in which the MLE is located has taxing rights over the service or intangible acquired by the MLE".[60] Guideline 3.4 suggests that "where the customer has establishments in more than one jurisdiction, the taxing rights accrue to the jurisdiction(s) where the establishment(s) using the service or intangible is (are) located". This much is obvious. The difficulty arises when trying to determine which establishment is using a service or intangible. The report identifies three broad categories of approaches which are currently being used by different jurisdictions: (a) the direct-use approach, which focuses on the establishment that uses the service or intangible; (b) the direct-delivery approach, which focuses on the establishment to which the service or intangible is delivered; and (c) the recharge method, which focuses on the establishment that uses the service or intangible as determined on the basis of internal recharge arrangements within the MLE, made in accordance with corporate tax, accounting or other regulatory requirements.[61]

Each of these methods has its own difficulties. The report notes that the 'direct use' approach may not be particularly helpful where it is unclear at the time of the agreement which customer establishment will actually use the service or intangible, or where the service or intangible will be used by different establishments in different jurisdictions.[62] The utility of the direct delivery approach is limited to where there is physical delivery, and hence it is of little assistance in the kind of services and intangibles presently being discussed. The 'recharge method' is more complicated, so it is necessary to explain this approach in detail to identify its problems. Under the 'recharge method', MLEs must "internally recharge the costs of externally acquired services or intangibles to their establishments that use

[60] OECD, (n 6) 45.
[61] ibid.
[62] ibid, 46.

these services or intangibles, as supported by internal recharge arrangements".[63] A two-step approach is suggested by the report:[64]

> The first step follows the business agreement between the external supplier and the MLE. The taxing rights over the supply to the MLE are allocated to the jurisdiction of the customer establishment that represents the MLE in the business agreement with the supplier. The second step is required when the service or intangible is used wholly or partially by one or more establishments other than the establishment that has represented the MLE in the agreement with the supplier. This second step follows the internal recharge made by the MLE for allocating the external cost of the service or intangible to the establishment or establishments using this service or intangible. This internal recharge is used as the basis for allocating the taxing rights over the service or intangible to the jurisdiction where these establishment(s) of use is (or are) located, by treating this internal recharge of the externally acquired service as within the scope of VAT.

This method does appear to deal with the issue of the supplier (and potentially the customer) not knowing which establishment(s) will use the service or intangible at the date of the agreement, but it is not without problems. The report acknowledges that, in some circumstances, certain services may be centrally acquired for the use of several establishments. In such cases, a detailed analysis of the use by each of establishment would be unduly burdensome so, in practice, approximations are often used.[65] It would, therefore, be necessary to establish a set of internationally agreed rules concerning acceptable methods of approximation for each type of service. Administrations will also have to consider the potentially dramatic increase in the number of internal transactions that have to be audited to prevent avoidance and evasion.[66] If not managed carefully, the recharge method could have the effect of dramatically increasing compliance and collection costs. This would be a very ambitious undertaking indeed.

Ultimately the OECD stops short of recommending a single approach. Instead it views that each approach will have its own benefits in different

[63] ibid, 47.
[64] ibid, 57.
[65] ibid, 60-61.
[66] ibid.

circumstances, and that different approaches can be combined when appropriate.[67] Given the above discussion, it seems that the recharge method is the most accurate way of determining destination in a VAT context. It thus comes as no surprise that it is favoured by Devereux and de la Feria.[68] A number of lingering issues have been mentioned but they are, at least theoretically, possible to resolve. Yet, it is submitted that it may not be possible to straightforwardly apply this to the DBCFT.

At first glance, it appears that there is no structural issue which prevents the recharge method being applied to the DBCFT. Like in a VAT context, following the internal recharge arrangements of a multinational allows tax authorities to track where services and intangibles end up being used, and hence where the 'true' destination lies. Nonetheless, it is argued that, on closer inspection, one can see that in the context of a direct tax, the recharge method is vulnerable to manipulation and could lead to significant avoidance. This can be seen by considering some variations of the following example: X sells an intangible to A, and the intangible is used in A's establishments in jurisdictions B and C. The internal recharge method allows authorities to see the real destination of the intangible, as well as how much of the cost each establishment is responsible for.

While this approach gives the multinationals the possibility to manipulate or misstate the internal recharge figures, in the context of VAT, there is no incentive to do so. The credit invoice method means that a business is charged VAT on each purchase it makes, but businesses are usually able to claim a credit or deduction for the VAT paid so long as the purchase was a business input. Hence, it is only the final consumer that actually pays the tax. This reflects the fact that VAT is a tax on consumption rather than general spending. Manipulating the figures to make it appear that the destination of an intangible is different from its actual location would, therefore, have no tax consequences for businesses. Yet, this may not be true in all situations under the DBCFT. To begin with, in the second version of the tax,[69] X pays no tax on the export, but the customer is unable to deduct the cost of the import from its tax base. As mentioned above, the tax operates under the subtraction-based method so businesses will be liable to pay the tax. As a consequence, if there is a way to manipulate the internal recharge mechanism to give the impression that the destination of an import is a low tax jurisdiction,

[67] ibid, 48. For example, the direct-use method is said to be particularly effective where the use of the service by one or more establishments is obvious.

[68] Devereux and de la Feria, (n 7) 17.

[69] This is the version where the customer pays no direct tax on the import but is unable to deduct the cost of the import from its tax base.

businesses may take such opportunity.[70] Under the second version of the tax, the non-deductibility of imports in all jurisdictions other than that of the seller means that an MLE is unable to shift its taxable profits in the same way it would under the source-based system.[71] Yet, it is possible to conceive of situations where manipulating the destination under the recharge method would result in more favourable treatment.

The first instance is where the multinational has an establishment in the same jurisdiction as the seller. If the internal recharge figures are manipulated to make it appear that the destination of the transaction is the jurisdiction of the seller, it would not be treated as an import. It would, therefore, be treated as deductible, resulting in lower tax liability. As mentioned, the OECD points out that many services, such as accountancy or legal services, are acquired by one establishment on behalf of the wider group; hence, it would not be difficult for an MLE to acquire such services in the same jurisdiction as the service provider and manipulate the recharge figures so that most, if not all, of the cost is attributed to the same jurisdiction. This problem would be exacerbated further if the service provider is also an MLE because the customer would then be able to choose the jurisdiction with the lowest tax rate. It is true that the provider may resist such an arrangement because, if the transaction is not characterised as an 'export', it is included in their tax base.

A few points can be made about this. Firstly, from a practical perspective, it is perfectly possible that in such a situation the customer may have the market power to insist on such an arrangement. In any case, it is likely that the formal destination of the services would become a bargaining chip in negotiations. This is surely undesirable. Secondly, the tax outcome in such a situation is heavily dependent on how each system treats different types of establishments. I will return to this point further below, but, for now, the problem can be demonstrated by using a simple example. In the situation where both parties are MLEs, it is possible that the establishments of each MLE may be characterised differently depending on exactly what form they take. To develop the example above, suppose that X has an establishment in jurisdiction B (B1) alongside A's office there (B2). If jurisdiction B did not, for whatever reason, recognise the existence of B2 but did recognise the existence of B1, the jurisdiction may view the transaction as being from jurisdiction B to jurisdiction C. This could be the case even if the purchaser's internal recharge figures suggest it is B2 which is using the service or intangible.

[70] As mentioned, tax authorities may seek to audit these figures, however, this will not lead to all manipulations being spotted. Further, where services are acquired by one establishment for the benefit of all businesses may legitimately locate the import of such services in one branch.

[71] Devereux and de la Feria, (n 7).

The consequence would be that jurisdiction B views the transaction as an export and so does not subject it to tax, but jurisdiction C does not view the profits of B2 as being subject to C's tax jurisdiction. The profits of B2, and the transaction between B1 and B2, would therefore go untaxed. Different results follow if it is B1 that is not recognised; jurisdiction B would view the transaction as being between A and B2 and so an import and non-deductible. The profits of B1 would not be taxed in either jurisdiction A or B, with the result that whilst B2 is taxed when it shouldn't be, B1 is not taxed when it should be.

If the first version of the DBCFT is adopted, the above analysis applies. Nevertheless, there is a more basic problem. In the original example, where only the customer is an MLE, manipulation of the internal recharge figures could straightforwardly be used to have the destination of the transaction in the most tax favourable jurisdiction.[72] Further, if the chosen destination was not the jurisdiction of the supplier, the supplier would have no reason to protest.

There is one further potential problem that is left unexplored by the literature on the DBCFT. While the authors say that either version of the tax could be adopted and that both versions are economically equivalent,[73] they do not say whether all jurisdictions must adopt the same version. If not, there is a further opportunity for avoidance as multinationals can alter their behaviour as appropriate. Suppose jurisdiction B chooses to implement the second version of the tax whereas jurisdiction C chooses to implement the first version. Branch B is an exporter with little taxable profits, and branch C sells primarily to the domestic market so has significant taxable profits. If a service is imported and used by both establishments, it may be advantageous to manipulate the figures so that the destination of the import appears to be jurisdiction B. As B's tax liability is connected to its profits, B will owe little tax. In jurisdiction C on the other hand, the same amount of tax will be levied, regardless of the branch's profits. If branch B has domestic sales of 50, labour costs of 20 and the import costs 40, its taxable profits will be 30. Assuming a tax rate of 10%, it will have a total tax liability of 3. If C has domestic sales of 120, labour costs of 20 and the import costs 40, it will have taxable profits of 60. Assuming a 10% tax on the import *and* the branch's profit, it will have a total tax liability of 10 (6+4). This may appear to be a convoluted route for a multinational to go down, and can be easily prevented by having all jurisdictions adopt the same version of the tax. Whether such consensus would actually be possible may come

[72] As stated above, this analysis is less likely to apply to businesses as it seems more likely that final consumers rather than businesses would face this version of the tax.

[73] Auerbach, Devereux, Keen, and Vella, (n 10) 29.

down to the preferences of each state and their existing tax structures. It certainly shouldn't be taken for granted that consensus could easily be achieved.

Hence, whilst the proponents of the DBCFT are right to argue that the tax eliminates the main methods of avoidance under the source-based system, this is not the whole picture. A lack of detailed discussion of the internal recharge mechanism, particularly in connection to branch mismatch arrangements (see Part II section C), means that the there is no recognition of the fact that this mechanism may not be suitable for the DBCFT.[74] It appears that formalism still exists in the proposed system, albeit regarding imports rather than profits.

2. Efficiency and Competition

Given these conclusions, the arguments outlined above in relation to efficiency and competition do not necessarily hold. In addition to providing an incentive to manipulate internal recharge figures, the DBCFT may also provide an incentive to structure transactions in particular ways to minimise tax liability. It flows that there might also be an incentive for countries to compete. It is true that different tax rates may not act as incentives in the same way as they would under the source-based system, but states may compete in other ways. For example, a state may choose to create favourable conditions for exporting establishments or may choose not to recognise the existence of certain branches. It, therefore, appears that the three core problems of formalism in a source-based system are still present under the DBCFT.

B. Branch mismatch arrangements

As mentioned, one of the key issues not discussed in any of the DBCFT literature is that of branch mismatch arrangements. The keen-eyed reader may have spotted that, whilst the OECD's VAT/GST report often makes reference to the 'establishments' of a multinational, there is surprisingly little detail on what this means. Footnote 24 says:

> For the purpose of these Guidelines, it is assumed that an establishment comprises a fixed place of business with a sufficient level of infrastructure in terms of people, systems and assets to be able to receive and/or make supplies. Registration for VAT purposes by itself does not constitute an establishment for the purposes of these Guidelines. Countries are encouraged to publicise what constitutes an "establishment" under their domestic VAT legislation.[75]

[74] Devereux and de la Feria, (n 7) 17; Auerbach, Devereux, Keen, and Vella, (n 10) 79.
[75] OECD, (n 6).

The working definition for the purposes of that report is of little importance in the present context. What is important, however, is the implicit acknowledgement that different countries may have different rules about what constitutes an 'establishment'. The problem will be laid out more fully below, but, in short, the issue is that jurisdictions often have different rules about how to characterize establishments in their territory resulting in a mismatch. Unfortunately, because of space limitations, this issue cannot be given the full attention it deserves here, so the rest of this paper will focus on the basics, and how this relates to the discussion about tax.

The OECD report on branch mismatch arrangements[76] came after a wider report on hybrid mismatches that addresses a range of issues including, inter alia, the characterisation of financial instruments.[77] It should be remembered that whilst the issue of branch mismatch arrangements is more obviously related to the destination principle, the following discussion may be just a small representative sample of the wider issues. Firstly, there is an overlap between the branch and hybrid mismatches because both can lead to the same problems of competition and efficiency, and many structures discussed in the branch mismatch report can also be found in the wider report on hybrid mismatches.[78] Nevertheless, branch mismatch arrangements remain distinctive. The OECD explains that, whilst hybrid mismatches generally are the product of inconsistencies because of different treatment of entities or financial instruments across jurisdictions; branch mismatches arise from the way the head office and the branch account for a payment made to or from a branch.[79] It therefore seems that taxpayers themselves can be just as much of a cause of the problem as the tax authorities, giving rise to avoidance and evasion opportunities.

There are three outcomes that may arise from branch (and hybrid) mismatch arrangements: (a) D/NI: "the payment is deductible under the rules of the payer jurisdiction but not included in the ordinary income of the payee"; (b) DD: "where the payment triggers two deductions in respect of the same payment"; and (c) Indirect D/NI: "where the income from a deductible payment is set off by the payee against a deduction under a hybrid mismatch arrangement".[80]

In total, there are five broad categories of arrangements that can lead to one of these outcomes. Unfortunately, space precludes discussion of all of these

[76] OECD, *Neutralising the Effects of Branch Mismatch Arrangements, Action 2: Inclusive Framework on BEPS* (OECD Publishing 2017).

[77] OECD, *Neutralising the Effects of Hybrid Mismatch Arrangements, Action 2 - 2015 Final Report* (OECD Publishing 2015).

[78] OECD, (n 76) 10-11.

[79] ibid, 9-10.

[80] OECD, (n 77).

structures, so the rest of this paper will focus on the first two. These arrangements have been chosen because they are the most easily applicable to the DBCFT. The five categories are the following:

a. Disregarded branch structure. Under this structure, the jurisdiction of the headquarters (the residence jurisdiction) treats a payment as being made to a foreign branch and exempt from income. The branch jurisdiction, however, does not recognize the existence of the branch and so does not collect any tax. There are a range of ways in which this could arise.[81]

b. Diverted branch payment. This payment has the same structure as a disregarded branch, but the mismatch arises because of a difference in laws for the attribution of payments. Hence, even where the branch jurisdiction recognises the existence of the branch, if each jurisdiction treats a payment as if made to the establishment in the other neither will levy any tax. The residence jurisdiction will treat the payment as if it was made to the branch and hence deductible from the head office's income, whilst the branch jurisdiction will treat the payment as if it was paid to the head office and hence not included in the branch's income.[82]

c. Deemed branch payment. This mechanism takes advantage of rules which allow taxpayers 'to recognise a deemed payment between two parts of the same taxpayer'. Where there is no corresponding adjustment in the net income of the two entities to take the effect of this payment into account, internal mismatches can arise. The OECD gives the following example to help demonstrate this arrangement: "A Co supplies services to an unrelated company (C Co) through a branch located in Country B. The services provided by the branch exploit the underlying intangibles owned by A Co. Country B attributes the ownership of those intangibles to the head office and treats the branch as making a corresponding arms-length payment to compensate A Co for the use of those intangibles. This deemed payment is deductible under Country B law but is not recognised under Country A law (because Country A attributes the ownership of the intangibles to the branch). Meanwhile, the services income received by the

[81] ibid, 14-15.
[82] ibid, 15-16.

branch is exempt from taxation under Country A law because of an exemption or exclusion for branch income in Country A".[83]

d. DD branch payments. Such payments occur when one payment results in deductions in two different jurisdictions. The OECD outlines two ways in which such mismatches may arise. The first is where the residence jurisdiction provides the head office with an exemption for branch income while permitting it to deduct the expenditures attributable to the branch. If the branch jurisdiction also allows the taxpayer to claim a deduction for the same expenditure, a mismatch can arise.[84] The second way in which DD branch payments can arise is in situations where the residence jurisdiction considers all the income and expenditure of a taxable branch for tax purposes.[85] The main way of this occurring is when "the branch is permitted to join a tax group or there is some other mechanism in place in the branch jurisdiction that allows expenditure or loss to be offset against income derived by another person that is not taxable under the laws of the residence jurisdiction".[86]

e. Imported branch mismatches. Here, the mismatch arises where an entity with a "deduction under a branch mismatch arrangement offsets that deduction against a taxable payment received from a third party.[87] This results in an indirect D/NI outcome.

Each of these structures allows taxpayers to manipulate the source principle by exploiting the different treatment of the principle across different jurisdictions. In particular, the first two arrangements give taxpayers the opportunity to hide part of their taxable profits by locating branches in strategic locations where the source principle treats payments as being made to the head office. We, therefore, seem to have the same problem of formalism seen elsewhere. Of course, taxpayers are not entirely to be blamed for these mismatches as they have no control over the tax laws of the states in which they operate but are also not completely innocent. Using the example of diverted branch payments, the OECD points out that the mismatch could be eliminated if the taxpayer was to take the relevant payment into

[83] ibid, 16-17.
[84] ibid, 17.
[85] ibid.
[86] ibid, 16-17. See the example on pages 17 and 18 of the Guidelines more details.
[87] ibid, 18-19.

account in either the residence or branch jurisdiction.[88] Businesses will naturally make the decision whether to open a branch in a foreign jurisdiction on a variety of commercial factors and it would be foolish to argue that taxpayers are actively and aggressively pursuing tax avoidance opportunities in this way. Nevertheless, given that there is empirical evidence[89] showing that tax can and does have some influence in the decision-making processes of businesses, we should acknowledge that the arrangements just described have the potential to be a problem. This is even more so if they apply in a DBCFT context, as the OECD's proposed solutions may not be suited to a destination-based tax.

Before considering the solutions, it is important to understand how each arrangement would work in a DBCFT context. Given that imports are the effective subject of the tax, the discussion must revolve around them rather than payments and profit more generally. Hence, with regard to the disregarded branch structure, we are considering the situation where a branch pays for an imported good or service, but the branch jurisdiction does not recognise the existence of the branch. This means that, under the second version of the DBCFT, its profits (from which there would be no deduction for the import) are not taxed, so there is no effective tax on the import. As the residence jurisdiction views the branch as a separate taxable entity, the tax authorities do not include the branch's profits in its calculation of the head office's profits, and so the import is not taxed there either. Under the first version of the DBCFT, the situation may be more complicated because, even though the branch jurisdiction would again not recognise the existence of the branch, it is unclear what the effect of this would be in the context of a tax levied directly.

With regard to services and intangibles, the problem discussed in the previous section would again be relevant. Assuming, however, that the internal recharge method is a workable solution in the DBCFT context, and that the destination of all products could be ascertained, the question changes on how imports should be defined. If imports are defined by reference to their physical location or where they are *actually* used, there would be no problem in identifying the jurisdiction which should tax the import. If, on the other hand, imports are defined by reference to where they are *legally* considered to be used, there may be more of a problem. As the branch jurisdiction (the jurisdiction of actual use) does not recognise the existence of the branch and attributes all its activities to the residence jurisdiction, it will not recognise the import and therefore would not tax it. Furthermore, if the residence jurisdiction does recognise the existence of the branch, it will surely view the import as being one to the branch jurisdiction and

[88] ibid, 10.

[89] Auerbach, Devereux, and Simpson, (n 5) 855.

will not tax the import either. The situation could be complicated further if the two jurisdictions take different approaches on the characterization of imports. This issue clearly needs to be considered further.

The problem of the diverted branch payment can also arise in the DBCFT context. Under the second version of the DBCFT, if the branch jurisdiction attributes a payment to the residence jurisdiction and vice versa, the taxable profits in both jurisdictions will be lower, meaning that the effective tax on any imports will also be lower. Under the first version of the tax, the businesses' profits are not related to the amount of tax payable and are instead levied on the value of the import. Hence, a mismatch in the attribution of branch payments will have no effect on tax liability. But, if we instead consider the possibility of 'diverted branch imports' where there is a mismatch in the rules for determining the destination of an import, a similar problem will arise.

It should be said from the outset that the solutions offered by the OECD are not perfect and will not eliminate the problem even in a source-based context. It is inevitable that different jurisdictions will have different rules, so the OECD's recommendations are merely an attempt to minimise the damage such mismatches can cause. Recommendation 1 of the OECD report suggests that:

> [J]urisdictions that provide an exemption for branch income should consider limiting the scope and operation of this exemption so that the effect of deemed payments, or payments that are disregarded, excluded or exempt from taxation under the laws of the branch jurisdiction are properly taken into account under the laws of the residence jurisdiction.[90]

This recommendation is intended to address all five mismatch arrangements, rather than just the two being described in detail here. Nevertheless, it must be the starting point for the present discussion as the OECD sees this as the default and preferable approach because of a number of advantages it has over the other recommendations. Such advantages include its applicability to a wide range of arrangements and the comparative ease with which the head office can identify the payment that gives rise to the mismatch.[91] The report points out that there are a number of ways in which this could be successfully implemented in a source-based system, including by "requiring that any payment, which is derived by a resident taxpayer and not subject to tax in the branch jurisdiction, be brought into charge

[90] OECD, (n 77) 24.
[91] ibid, 24 and 58.

to taxation in the head office [and] limiting the branch exemption to the amount of net income actually brought into the charge to tax by the branch".[92]

Recommendation 2 of the OECD report deals with the issues of branch mismatch arrangements and diverted branch payments.[93] As the two issues are very similar in structure, the following discussion applies to both, unless stated otherwise. In respect of these two mechanisms, the OECD suggests that the "payer jurisdiction should deny a deduction for a payment that gives rise to a D/NI outcome"[94] to the extent that the outcome is the result of one of these two mechanisms.

As mentioned in Part II section B(1), the OECD admits from the outset that these recommendations, as well as others that are contained in the report, are unlikely to provide a complete solution to the formalism of the system, even where the tax is based on 'source' rather than on 'destination'. Proponents of the DBCFT must show either that the solutions will have at least some effect where corporations are taxed in accordance with the destination principle, or that alternative solutions exist. Space prevents full examination of the issue here, but prima facie it appears that there may be some problems in straightforwardly applying these solutions to the DBCFT. Such difficulty is simply because these solutions focus on payments rather than imports. Whilst it may be possible to adapt the OECD guidelines, a number of issues would have to be resolved, especially considering the discussion above on MLEs. Would the residence jurisdiction view the transfer as an import, despite the fact that the branch jurisdiction would not view it as an export? Would the headquarters have to pay an arms-length price, even though the branch jurisdiction does not recognise the existence of any 'external'[95] transaction? More generally, how are the headquarter to be identified, and what mechanisms could be put in place to minimise avoidance opportunities? How would any solutions prevent the manipulation of destination where the transaction involved services or intangibles? All these questions may have answers, but they should be fully explained and debated. As far as the author is aware, Devereux and others have not yet explored the problem of branch mismatch arrangements, let alone the wider issue of hybrid mismatch arrangements. We are therefore a long way from having the answers that we need.

III. Conclusion

The formalism in the current sourced-based corporate tax system has led to a number of problems which are clear for all to see. The proponents of the

[92] ibid.
[93] ibid, 27.
[94] ibid, 27.
[95] In example, transactions between two independent persons.

DBCFT put forward a strong case that their model will eliminate many of the problems faced by the source-based system. Despite their proclamations, much of their analysis has come from an economic rather than a legal perspective. This paper has sought to reveal the formalistic aspects through a legal analysis of the DBCFT by exploring just two of the problems related to destination which have received little if any attention in the literature discussing the tax. Further, it is far from obvious that any of the existing solutions to these problems can be implemented effectively in a DBCFT context. Others may still wish to argue that, despite these problems, the DBCFT is still preferable to the current source-based approach. The author does not wish to make any comment on this, but it is clear that, if the DBCFT was ever to be implemented, the formalism described above and its associated problems would have to be addressed.

Evaluating the Need To Reform Northern Ireland's Abortion Law From a Human Rights Perspective

*Zoe Louise Tongue**

Introduction

Northern Ireland has one of the most restrictive abortion regimes in Europe and is the only part of the United Kingdom where access to abortion is almost entirely illegal. On the 25th of May 2018, Ireland voted overwhelmingly in favour of repealing the Eighth Amendment in a referendum, and the House of Keys recently passed the Abortion Reform Bill which decriminalises abortion on the Isle of Man, leaving Northern Ireland the only part of Britain and Ireland where women[1] have virtually no access to abortion.

This article argues that Northern Ireland's restrictive abortion law violates the UK's human rights obligations, and it is therefore necessary to provide for abortion on the grounds of rape and fatal foetal abnormality (FFA) and to fully decriminalise abortion. Section I highlights the comparative restrictiveness of Northern Ireland's abortion law, and how the failure to liberalise the law is the result of the dominant morally conservative, anti-abortion attitudes of political and religious figures. This public discourse, coupled with the draconian criminal offence, creates an environment where women cannot access abortions even on lawful grounds, and this forces abortion-seeking women to travel abroad or illegally terminate their own pregnancies.

Section II considers the human rights arguments for reforming Northern Ireland's abortion law, focusing chiefly on the right to freedom from inhuman and degrading treatment under Article 3 of the European Convention on Human

* LLB, LLM (Warwick). I'd like to thank Dr Aisling McMahon for her help with this piece.

[1] Where the term 'woman' is used, it is acknowledged that trans* men and non-binary people also require access to abortion services.

Rights (ECHR) and Article 7 of the International Covenant on Civil and Political Rights (ICCPR). The right to privacy under Article 8 ECHR is also considered insofar as States have a positive obligation to ensure that lawful abortions are accessible in practice. It is argued that Northern Ireland's prohibition on abortion on the grounds of rape and FFA is a violation of Article 3 and, extending this argument, that the criminalisation of abortion also amounts to inhuman and degrading treatment. Section 3 examines how Northern Ireland's abortion law should be reformed to guarantee access to abortion in cases of rape and FFA, arguing that the Abortion Act 1967 should be extended or abortion on request be provided for, and it is concluded that decriminalisation is also required alongside non-legal measures to ensure that abortion services are accessible in practice.

I. The Restrictiveness of Northern Ireland's Abortion Law

Abortion is legal in almost every European state; Andorra and Malta are the only European countries with abortion laws more restrictive than Northern Ireland. Unsafe, illegal abortions have been increasingly recognised as a consequence of restrictive abortion laws, and it is estimated that 30 women die for every 100,000 unsafe abortions in developed countries.[2] The Abortion Act 1967 was passed in England, Wales, and Scotland following the acknowledgment of backstreet abortions as a social problem which resulted in a significant maternal morbidity rate of around 50 women per year.[3] Since the 1967 Act was passed, there have been no recorded maternal deaths in England, Wales, and Scotland due to illegal abortions.[4] Yet, the Abortion Act has never been extended to Northern Ireland. This section will outline the restrictiveness of the Northern Irish position in comparison to other jurisdictions, and how reluctance to change the law and the stigmatisation of abortion is exacerbated by the dominant morally conservative public discourse and prejudicial attitudes which fail to reflect the views of the majority of Northern Ireland's population. This restrictive legal position has substantial negative impact, creating a chilling effect on medical practitioners and

[2] World Health Organisation (WHO), 'Preventing unsafe abortion' (*World Health Organisation*) <www.who.int/mediacentre/factsheets/fs388/en/> accessed 17 February 2018.

[3] Wendy Savage, 'Fifty years on, the Abortion Act should be celebrated – and updated' *The Guardian* (London, 27 October 2017) <www.theguardian.com/commentisfree/2017/oct/27/50-years-abortion-act-law-women> accessed 17 February 2018.

[4] Katherine Side, 'Contract, Charity, and Honourable Entitlement: Social Citizenship and the 1967 Abortion Act in Northern Ireland after the Good Friday Agreement' (2006) 13(1) *Social Politics* 89, 98.

forcing women to travel abroad for abortions or to terminate their pregnancies illegally, risking prosecution and their own health.

A. The comparative restrictiveness of northern ireland's abortion law

Northern Ireland's abortion law is guided primarily by the Offences Against the Person Act 1861, which provides for the offence of unlawfully administering or procuring drugs or using instruments to cause a miscarriage with a maximum penalty of a life sentence.[5] Section 25 of the Criminal Justice Act (NI) 1945, which mirrors the Infant Life (Preservation) Act 1929, created the offence of child destruction, again with the penalty of a maximum life sentence, but also introduced a good faith exception for the purpose of preserving the life of the woman. The application of the Infant Life (Preservation) Act exception was clarified in *Bourne,* a case involving a doctor who performed an abortion for a 14-year-old girl after she became pregnant from rape.[6] Macnaghten J found that the word "unlawfully" in section 58 of the 1861 Act could be read as importing the good faith exception found in the Infant Life (Preservation) Act.[7] Thus, *Bourne* created an exception to the OAPA allowing abortion for the purpose of preserving the life of the woman, which covered circumstances where the woman would become "a physical or mental wreck".[8] This test was found to be satisfied by the risk of suicide in two Northern Irish cases, where the court ordered in each case, both concerning minors, that abortions could be lawfully obtained.[9] The Northern Irish courts have, however, subsequently limited the scope of the *Bourne* exception. In *NHSSB v A*, MacDermott J emphasised that any adverse effect on the pregnant woman's physical or mental health must be "real and serious" and that it is a question of fact and degree as to whether the effect is "sufficiently grave to warrant terminating the unborn child".[10] The meaning of "real and serious adverse effect" was further restricted to one which is "permanent or long-term", rather than short-term, by Pringle J in *WHSSB v CMB.*[11] Therefore, the only grounds for a lawful abortion in

5 Offences Against the Person Act 1861, ss 58 and 59.
6 *R v Bourne* [1939] 1 KB 687 (Court of Criminal Appeal).
7 ibid 691.
8 ibid 694.
9 *Northern Health and Social Services Board v F and G* [1993] NI 268 (NIFam); *Re CH (a minor)* (NI High Ct, 1995).
10 *Northern Health and Social Services Board v A and others* [1994] NIJB 1, 5.
11 *Western Health and Social Services Board v CMB and the Official Solicitor* (NI High Ct, 1995).

Northern Ireland are when the woman's life is at risk, including a risk of suicide, or there is a serious risk to her long-term health.

The restrictiveness of this position is demonstrable by comparison to the Abortion Act, which sets out a range of broader exceptions to the OAPA. Under the 1967 Act, a person shall not be guilty of an offence if a pregnancy is terminated by a medical practitioner, provided that two medical practitioners are satisfied that the abortion would fall under one of four grounds: where continuing the pregnancy poses a greater risk to the physical or mental health of the woman than an abortion, provided that the pregnancy has not exceeded 24-weeks; to prevent grave permanent injury to the physical or mental health of the woman; if there is a risk to the life of the woman; or if the foetus would be born with a serious disability.[12] The first exception, the social ground, effectively allows any woman to have an abortion, given that continuing a pregnancy poses greater risks to her physical and mental wellbeing than an abortion.[13] The position in many jurisdictions around Europe goes further, allowing abortion on request (requiring no specific circumstance) up to the end of the twelfth week of pregnancy.[14] Even in countries with highly restrictive abortion laws, such as Poland, provision is made for abortions on the grounds of rape and FFA, as these are generally regarded as exceptional circumstances alongside risk to the life of the pregnant woman. Northern Ireland is therefore an outlier from the general European consensus on access to abortion, even by restrictive standards.

Furthermore, Malta and Andorra are the only two European states with more restrictive abortion laws than Northern Ireland, and yet they all have lower maximum sentences for the criminal offence of abortion, highlighting the draconian nature of the outdated 1861 Act. Andorra also only permits abortion to save the pregnant woman's life, with a penalty of two-and-a-half-years for a woman who terminates her own pregnancy,[15] while Malta, which prohibits abortion in all circumstances, imposes a lower sentence on of a maximum of three years for the abortion-seeking woman.[16] Before the Eighth Amendment was repealed, Ireland only allowed abortion where there was a "real and substantial risk of loss of the woman's life", including the risk of suicide,[17] but excluding any

[12] Abortion Act 1967, section 1.

[13] British Medical Association, 'Decriminalisation of abortion: a discussion paper from the BMA' (British Medical Association 2017), 27–28.

[14] For example, Austria, Denmark, and Switzerland. See The Law Library of Congress, *Abortion Legislation in Europe* (The Law Library of Congress 2015).

[15] UN Department of Economic and Social Affairs (Population Division), 'Abortion Policies: A Global Review' (2002).

[16] Criminal Code of Malta, section 241.

[17] Protection of Life During Pregnancy Act 2013, sections 7 and 9.

risk to the woman's health. The Protection of Life During Pregnancy Act 2013 repealed the relevant provisions of the OAPA and replaced them with the offence of intentionally destroying unborn life, which carried a maximum penalty of 14 years.[18] While Catherine O'Rourke suggested that abortion law in Northern Ireland was "marginally more liberal" than Ireland, the existence of a maximum life sentence made the Northern Irish position arguably harsher.[19] Thus, not only does Northern Ireland have one of the most restrictive abortion regimes in Europe, but it also imposes one of the most severe criminal sanctions.

B. Rejecting the claim of universal moral values in abortion context

The Northern Irish position has been borne out of the prominent moral conservatism of politicians and the central role of religious institutions in the abortion debate. This has been identified as a key barrier to abortion law reform, with the alignment of politics and religion evident in the lobbying of MPs by Northern Irish churches to prevent the extension of the Abortion Act and express condemnation for the opening of the Marie Stopes clinic in Belfast in 2012.[20] Fegan and Rebouché have criticised anti-abortion figures for their tendency to unquestioningly believe in the universality of their opinion.[21]This claim of near-unanimous public opinion has also been used as a continuous justification for the lack of action by Westminster, as abortion is treated as an issue of devolution which allows Parliament to abscond judgement on a controversial topic.[22] Recent polls refute this claim, showing that the majority of Northern Ireland does, in fact, support greater access to abortion: Amnesty International found that 58% of people thought that abortion should be decriminalised;[23] a poll conducted by Ulster University found that only 43% of people though that it should be illegal for a

[18] ibid section 22.

[19] Catherine O'Rourke, 'Advocating Abortion Rights in Northern Ireland: Local and Global Tensions' (2016) 25(6) *Social and Legal Studies* 716-740, 719.

[20] Fiona Bloomer and Kellie O'Dowd, 'Restricted access to abortion in the Republic of Ireland and Northern Ireland: exploring abortion tourism and barriers to legal reform' (2014) 16:4 *Culture, Health & Sexuality* 366, 367–368.

[21] Eileen Fegan and Rachel Rebouché, 'Northern Ireland's Abortion Law: The Morality of Silence and the Censure of Agency' (2003) 11 *Feminist Legal Studies* 221, 232.

[22] Jennifer Thomson, 'Explaining gender equality difference in a devolved system: The case of abortion law in Northern Ireland' (2016) 11 *British Politics* 371, 372.

[23] Amnesty International UK, 'Northern Ireland: Nearly 3/4 of public support abortion law change – new poll' (*Amnesty International UK*, 18 October 2016) <www.amnesty.org.uk/press-releases/northern-ireland-nearly-34-public-support-abortion-law-change-new-poll-0> accessed 17 February 2018.

woman to have an abortion because she does not want children;[24] and the Economic and Social Research council found that just 29.1% of people opposed changing Northern Ireland's abortion law.[25] The figures in support of decriminalisation are much higher for cases of rape (72%) and fatal foetal abnormality (67%).[26]

However, the domination by anti-abortion views of the mainstream discourse in Northern Ireland has resulted in prejudicial attitudes that focus on foetal life to the exclusion of the rights of pregnant women and heavy stigmatisation of abortion. This was highlighted in one Northern Ireland Assembly debate on preventing the extension of the Abortion Act, in which Britain's abortion rate was compared to the Holocaust.[27] Anti-abortion views are openly expressed by public figures; Jim Wells, just before becoming the NI Health Minister, said during a radio interview that abortion should not be available in cases of rape, arguing that the unborn child is the "ultimate victim".[28] The dominant anti-choice rhetoric expressed by religious and political figures has produced a stagnant area of law that has failed to evolve in line with the modern views of the population. Northern Ireland's restrictive regime is anomalous by European standards and ignores the consensus in Northern Ireland in support of liberalised access to abortion. There is no justification for preserving this outdated position.

C. The chilling effect of the criminal law

The restrictiveness of Northern Ireland's abortion law and the dominance of conservative values is a significant problem, as it creates a chilling effect on doctors who may be reluctant to perform even lawful abortions out of fear of prosecution. This chilling effect is exacerbated by the uncertainty of the law, which is based on a Victorian statute and only a few cases. In 2003, the Family Planning Association of Northern Ireland (FPANI) challenged the Department of Health, Social Services, and Public Safety for its failure to issue clear guidance as to the availability of abortion in Northern Ireland. [29] The Court of Appeal held that the Department had acted unlawfully in failing to provide advice to pregnant women

[24] Ulster University, 'Ulster University research reveals attitudes to abortion in Northern Ireland' (*Ulster University*, June 2017) <www.ulster.ac.uk/news/2017/june/ulster-university-research-reveals-attitudes-to-abortion-in-northern-ireland> accessed 17 February 2018.

[25] Jonathan Tonge, *Northern Ireland General Election Survey 2017* (*UK Data Service*, 2017) accessed 22 Mar 2019

[26] Amnesty International UK (n 23).

[27] NI Assembly Deb 20 June 2000.

[28] Adrian Rutherford, 'DUP's Jim Wells: Abortion should be ruled out for rape victims' *Belfast Telegraph* (Belfast, 25 August 2012) <www.belfasttelegraph.co.uk/news/northern-ireland/dups-jim-wells-abortion-should-be-ruled-out-for-rape-victims-28785234.html> accessed 07 January 2018.

[29] *Family Planning Association of Northern Ireland v Minister for Health, Social Services, and Public Safety* [2004] NICA 37–39, [2005] NI 188.

and medical professionals. Nicholson LJ found that medical practitioners were not adequately aware of the law relating to abortion, and the Department should have recognised this following the refusal of clinicians to carry out a court-ordered termination due to the existence of the criminal offence.[30] He further held that the Department had a duty to ensure that accurate guidance is given, pointing out that it was not good enough for the principles governing the law to be found only in case law, of which two of the four key cases are unreported.[31]

Guidance was not issued until five years after this case, and was clearly not intended to remedy the concerns raised in *FPANI* decision given that it perpetuates the chilling effect of the law by reinforcing the illegality of abortion in Northern Ireland and restricts the law further by recommending the certification of two medical practitioners as good practice,[32] which has no basis in Northern Irish law. Despite this conservative position, it was still immediately challenged by the Society for the Protection of Unborn Children on the grounds that it failed to acknowledge the presumptive illegality of abortion in Northern Ireland and the rights of the unborn.[33] Girvan J gave a misdirection as to the law in his judgment, asking: "Could the giving of such advice constitute an offence of counselling or procuring an abortion unlawful in Northern Ireland?"[34] Not only was the answer to be found in the negative, as non-directive counselling was already permitted in Northern Ireland and provided by FPANI,[35] but the judgment sustained the fear of medical practitioners that giving *any* advice to patients regarding abortion could put them at risk of prosecution. The most recent guidance does provide for counselling services [36] and it was also recently clarified that staff in Northern Ireland would not be at threat of prosecution for referring women to abortion clinics in England and Wales.[37]

However, this guidance remains cautionary, clearly emphasising that abortion is a criminal offence, and the chilling effect therefore remains. This effect

30 ibid [93].

31 ibid [115].

32 Department of Health, Social Services, and Public Safety (DHSSPS), *Guidance on the Termination of Pregnancy: The Law and Clinical Practice in Northern Ireland* (2009) [1.8]

33 *Society for the Protection of Unborn Children, Re Judicial Review* [2009] NIQB 92.

34 ibid [37].

35 Family Planning Association (FPA), 'Pregnancy choices counselling service, Northern Ireland' (*Family Planning Association*) <www.fpa.org.uk/unplanned-pregnancy-and-abortion/unplanned-pregnancy-post-abortion-counselling-northern-ireland> accessed 19 March 2018.

36 DHSSPS, *Guidance for Health and Social Care Professionals on Termination of Pregnancy in Northern Ireland* (2016) pt 5.

37 Amelia Gentleman, 'No prosecution risk for Northern Ireland medical staff over abortion referrals' *The Guardian* (London, 7 September 2017) <www.theguardian.com/uk-news/2017/sep/07/no-prosecution-risk-for-northern-ireland-medical-staff-over-abortion-referrals> accessed 18 February 2018.

does not operate only in the context of giving advice; Bloomer and Fegan note that the threat of prosecution, coupled with the minimal training that doctors receive on abortion, means that doctors are unwilling to carry out the procedure even to save a woman's life.[38] The insufficient training on abortion in Northern Ireland is evidenced by students seeking education from external providers, such as Medical Students for Choice,[39] and Irish students travelling to London for abortion training.[40] The assertion that doctors will not perform life-saving abortions can be made based on the extremely low abortion statistics: in 2017/18, just twelve legal abortions were carried out in Northern Ireland.[41] In England and Wales, the majority of abortions are carried out under the social ground, so it is it is difficult to pinpoint exactly how many life-saving abortions are carried out each year, but it is likely to be significantly higher than twelve based on the figure of 197,533 abortions in total in 2017.[42] Thus, not only is abortion criminalised in almost every situation, but women who fall within the narrow exception and are legally entitled to an abortion may not be able to access one in practice.

D. Abortion tourism and illegal abortions: harming pregnant women

Women who require an abortion on any ground are therefore forced to either travel abroad to access abortion services or illegally terminate their pregnancies. In 2017, 919 Northern Irish women travelled to England or Wales for terminations.[43] This figure is likely to be an underestimate, as many women will withhold their Northern Irish addresses. Until recently, Northern Irish women were unable to access abortion services on the NHS in England and Wales but had to pay for their terminations at private clinics, despite being UK taxpayers. It was estimated that the cost of an abortion was between £200 and £2000, including travel and accommodation, placing a significant burden on women of low income who would have to borrow money in order to make this journey.[44] In 2017, the Supreme Court found that there was no duty for abortion services to be provided free of charge

[38] Fiona Bloomer and Eileen Fegan, 'Critiquing recent abortion law and policy in Northern Ireland' (2013) 34(1) *Critical Social Policy* 109, 113.

[39] Medical Students for Choice, 'MSFC Ireland' (*Medical Students for Choice*) <www.msfc.org/about-us/where-we-are/ireland/> accessed 18 February 2018.

[40] Catherine Shanahan, 'Medical students go to London for abortion training' *The Irish Examiner* (Cork, 10 June 2015) <www.irishexaminer.com/ireland/medical-students-go-to-london-for-abortion-training-335569.html> accessed 18 February 2018.

[41] Department of Health NI, *Northern Ireland Termination of Pregnancy Statistics 2017/18* (2019).

[42] Department of Health, *Abortion Statistics, England and Wales: 2017* (2018).

[43] ibid.

[44] Bloomer and Fegan (n 38) 111.

for Northern Irish women;[45] however, following this judgment an amendment was tabled by Labour MP Stella Creasy, supported by over 50 MPs, to allow Northern Irish women to access NHS abortions in England, leading the Health Secretary to concede and allow access to free abortion services.[46] The Scottish government later announced that it would do the same.[47] The Minister for Equalities and Women has since provided that Northern Irish women facing hardship will be eligible for grants to cover travel costs, and a central booking service is to be established in order to make accessing abortion services in England simpler.[48]

While this takes away the financial burden for a lot of women travelling to England for abortions, it does not help the most vulnerable women who may still be unable to afford to travel to England. Travelling requires taking time off work, resulting in a loss of pay, women who already have children may have to pay for childcare while they are away, and it remains unclear whether the hardship grants cover accommodation. Northern Ireland has a high poverty rate; in 2016/17, there were 336,000 people (18% of the population) living in relative poverty.[49] Thus, there will likely be a large number of women who cannot afford this option, even with the availability of funding; women with low incomes, homeless women, asylum seekers, young girls who are not financially independent, and those in abusive relationships thus remain unable to access safe, legal abortion services. Those able to travel still face the emotional burden of having to leave their families and support network behind, and the chilling effect on medical professionals means that they may not be able to access counselling or aftercare services upon their return.

Women unable to travel may be forced to have an abortion illegally in Northern Ireland, most often by taking the 'abortion pill' which induces a miscarriage in early pregnancies. The pills can be obtained relatively easily from online websites such as Women on Web and Women Help Women, which also provide help and advice for women seeking to terminate their pregnancies.[50] It has been estimated that there are around 3,000 discreet contacts to these providers each year from the whole of Ireland, with around a third of this number from

[45] *R (on the application of A and B) v Secretary of State for Health* [2017] UKSC 41, [2017] 1 WLR 2492.

[46] BBC News, 'Northern Ireland women to get free abortions in England' *BBC News* (London, 29 June 2017) <www.bbc.co.uk/news/uk-politics-40438390> accessed 06 January 2018.

[47] BBC News, 'Scotland offers free abortions to women from Northern Ireland' *BBC News* (London, 06 November 2017) <www.bbc.co.uk/news/uk-scotland-41879520> accessed 06 January 2018.

[48] HC Deb 23 October 2017, vol 630, col 192WS.

[49] Department for Communities (NI), *Poverty Bulletin: Northern Ireland 2016–17* (2018).

[50] Women on Web, 'Women on Web' (*Women on Web*) <www.womenonweb.org/> accessed 07 January 2018; Women Help Women, 'Women Help Women' (*Women Help Women*) <https://womenhelp.org/> accessed 07 January 2018.

Northern Irish women.[51] The websites supply Misoprostol and Mifepristone, two tablets that are on the World Health Organisation's list of essential medicines[52] and are thus considered to be relatively safe. However, women may seek abortion pills or other means of terminating their pregnancies from unsafe backstreet providers, and there is always the risk that a self-administered abortion could go wrong. Since 1967, there have been five deaths in Northern Ireland associated with complications from illegal abortions,[53] but this figure is likely higher assuming there are cases where it was not identified that a woman had an illegal abortion. Furthermore, where an illegal abortion does go wrong, women may be deterred from seeking medical attention. Sally Sheldon argues that there is "at least a hypothetical risk that heightened stigma regarding illegal abortion and fear of prosecution might lead to a reluctance to seek necessary aftercare", as a woman who confides in her doctor could be reported.[54] There have been a few recent cases where women have been prosecuted for terminating their unwanted pregnancies illegally. A three-month suspended sentence was given to a woman who induced a miscarriage using pills when she was nineteen, after she was reported to the police by her housemates,[55] and a mother faced prosecution for buying abortion pills to help her fifteen-year-old daughter.[56] Treating women as criminals for terminating their pregnancies has obvious repercussions for women's health, and clearly does not prevent abortions from taking place.

E. Conclusion

Section I has outlined the comparative restrictiveness of Northern Ireland's abortion law, manifested in the narrow exceptions for lawful abortions and the severity of the criminal offence. The claims of public figures as to the universal moral views of Northern Ireland create a barrier to reform and stigmatise women who obtain abortions, despite evidence that the majority of the population are in favour of more liberal abortion laws. The OAPA creates a chilling effect on doctors, who may be unwilling to provide advice or even perform live-saving abortions. The

[51] Sally Sheldon, 'How can a state control swallowing? The home use of abortion pills in Ireland' (2016) 24(48) *Reproductive Health Matters* 90, 92.

[52] WHO, 'WHO Model List of Essential Medicines 20th List' (2017).

[53] Side (n 4) 98.

[54] Sheldon (n 51) 94.

[55] Henry McDonald, 'Northern Irish woman given suspended sentence over self-induced abortion' *The Guardian* (London, 4 April 2016) < www.theguardian.com/uk-news/2016/apr/04/northern-irish-woman-suspended-sentence-self-induced-abortion> accessed 15 November 2017.

[56] Amelia Gentleman, 'Woman who bought abortion pills for daughter can challenge prosecution' *The Guardian* (London, 26 January 2017) < www.theguardian.com/world/2017/jan/26/ulster-woman-who-bought-abortion-pills-for-daughter-can-challenge-prosecution> accessed 15 November 2017.

only real options available to women seeking abortions are therefore to either travel abroad, imposing a financial and emotional burden, or have an unsafe, illegal termination. The impacts of restrictive abortion laws have become increasingly recognised as human rights violations, which will be addressed in the following Section as part of the argument for reform.

II. Northern Ireland's Abortion Law Violates Human Rights

While international human rights bodies are yet to affirm the existence of a right to abortion, it has been increasingly recognised by human rights bodies that restrictive abortion laws and the inaccessibility of safe, legal abortion can amount to human rights violations.[57] The extent to which the right to an abortion has been protected through human rights law is limited to exceptional circumstances, namely where continuing the pregnancy poses a risk to the woman's life or health, where the pregnancy is a result of rape, or in cases of FFA.[58] Thus, while women in Northern Ireland do not have a right to abortion *per se*, the region's restrictive abortion law which forces women to continue pregnancies in circumstances of rape and FFA could be found to violate human rights by constituting "tremendous cruelty".[59] The UK is a member state of the ECHR and signatory to the ICCPR and the Convention on the Elimination of Discrimination Against Women (CEDAW), and is required to comply with human rights standards established under these treaties. The UK is in violation of these standards due to the Northern Ireland's restrictive abortion law, and reform is therefore necessary.

While the ECtHR has been reluctant to find that Article 8 confers a right to abortion, the Court has found that abortion on grounds which are provided for in domestic law must be accessible in practice. The right to freedom from inhuman and degrading treatment under Article 3 of the ECHR and Article 7 of the ICCPR, as an absolute right, provides a stronger argument based on the adverse impacts caused to women by restrictive abortion laws. Both rights are engaged by the impact of prohibiting abortion in cases of rape and FFA, and the ECtHR and the Human Rights Committee (HRC) have been willing to find violations where abortion has been restricted on these grounds. This Section will also argue that the full decriminalisation of abortion is necessary to avoid violating Article 3 ECHR and Article 7 ICCPR. The full decriminalisation of abortion refers to the removal

[57] Johanna Fine, Katherine Mayall, and Lilian Sepulveda, 'The Role of International Human Rights Norms in the Liberalization of Abortion Laws Globally' (2017) 19(1) *Health & HRJ* 69, 71.

[58] Christina Zampas and Jaime Gher, 'Abortion as a Human Right – International and Regional Standards' (2008) 8(2) *HRL Rev* 249, 255.

[59] Chiara Cosentino, 'Safe and Legal Abortion: An Emerging Human Right? The Long-lasting Dispute with State Sovereignty in ECHR Jurisprudence' (2015) 15 *HRL Rev* 569, 583.

of the criminal offence of abortion on all grounds, as opposed to the legalisation of abortion in limited circumstances, a position which has been advocated by the CEDAW Committee.

A. Chilling effect of the OAPA violates Article 8

The ECtHR has recognised that Article 8 is applicable in cases involving abortion, but its willingness to find restrictive abortion legislation to be in violation of Article 8 has been confined to situations where the pregnant woman should have been allowed an abortion in domestic law but was arbitrarily denied one. Thus, while the UK could be found in violation of Article 8 for failing to ensure that Northern Irish women can access abortion services under the existing exceptions, Article 8 cannot be used to argue for more liberal abortion access because of the margin of appreciation. The ECtHR found a breach of Article 8 in *Tysiąc v Poland,* where a visually impaired woman was denied an abortion despite the existence of a health ground for abortion in Poland and confirmation by doctors that continuing her pregnancy could severely impact her vision.[60] In *P and S v Poland*, the ECtHR unanimously found a breach of Article 8 where a teenage girl should have been allowed a lawful abortion on the grounds of rape, stating that Poland had a "positive obligation to create a procedural framework enabling a pregnant woman to effectively exercise her right of access to lawful abortion."[61] In *R.R. v Poland*, a violation was found where the applicant was denied genetic tests until the time limit for a legal abortion based on foetal abnormality had expired.[62]

ABC v Ireland concerned three applicants, all of whom became pregnant unintentionally and travelled to England for abortions.[63] The first applicant felt unable to continue her pregnancy due to her history of depression and alcoholism, and the second applicant had taken the morning-after pill, which failed. The ECtHR cited the "profound moral values of the Irish people"[64] in finding that the restriction on abortion pursued a legitimate aim and afforded a wide margin of appreciation to Ireland in finding that a fair balance was struck between the Article 8 rights of the applicants and the protection of the unborn, so no breach was found.[65] However, the third applicant had previously undergone three years of chemotherapy and her cancer was in remission, but feared that the pregnancy would be life-threatening if her cancer returned, as she would then be unable to

[60] *Tysiąc v Poland* [2007] ECHR 219.
[61] *P and S v Poland* [2012] ECHR 1853.
[62] *R.R. v Poland* [2011] ECHR 828.
[63] *A, B, C v Ireland* [2010] ECHR 2032.
[64] ibid [230].
[65] ibid [235]–[241].

have treatment in Ireland, and she was given insufficient information as to the risks. The Court found a breach of the third applicant's Article 8 rights as Ireland had failed to comply with its positive obligation to implement an accessible and effective procedure to establish whether she qualified for a lawful abortion or not.[66] The Court also acknowledged that the criminal provisions of the OAPA 1861 coupled with the substantial uncertainty of the law constituted a serious chilling effect for women and doctors.[67]

This judgment has been criticised for its "remarkably value-free judgements, avoiding references to female autonomy or dignity, or to discrimination against women"[68] and for placing abortion "firmly within the domestic sphere and [leaving] apparently complete discretion to the State."[69] There is an argument to be made that the ECtHR incorrectly focused on the moral views of Ireland and the lack of a European consensus on the beginning of life, rather than on the existence of a consensus regarding access to abortion, but the application of the margin of appreciation makes it unlikely that the Court will find a violation of Article 8 in circumstances where abortion is not already provided for in domestic law. However, *ABC*, alongside the Polish cases, is significant in confirming that Article 8 confers a positive obligation on States to ensure that legal abortions are accessible. As established above, the chilling effect of the OAPA leaves Northern Irish doctors reluctant to perform abortions on the lawful grounds. Thus, the UK is in breach of its positive obligation to ensure that women can access abortions based on the exceptions already recognised in Northern Ireland's domestic law, and reform, at the very least, is required by Article 8 to remedy this.

B. Article 3 and rape

It is necessary for the UK to ensure that Northern Irish women can access abortion on the grounds of rape, as the current prohibition on abortion on this ground may amount to a violation of Article 3. In *P and S*, a breach of Article 3 was found due to the horrific treatment and lack of protection the applicant, a victim of sexual abuse, had received; she was manipulated by doctors when she attempted to access legal abortion services, harassed by anti-abortion activists, and her mother was accused of forcing her to terminate the pregnancy.[70] With the caveat that rape

[66] ibid [267].

[67] ibid [254].

[68] Daniel Fenwick, 'Abortion Jurisprudence' at Strasbourg: deferential, avoidant, and normatively neutral?' (2014) 34 *Legal Studies* 214, 229.

[69] Fiona De Londras and Kanstantsin Dzehtsiarou, 'Grand Chamber of the European Court of Human Rights, *A, B & C v Ireland*, Decision of 17 December 2010' (2013) 62 *Intl & Comp LQ* 250,, 261.

[70] *P and S* (n 61).

was already a ground for abortion in Poland, this case demonstrates the willingness of the ECtHR to recognise that the distress caused by being unable to obtain an abortion is sufficiently severe that it amounts to an Article 3 violation. Given that Article 3 is an absolute right, the fact that abortion is not already lawful on the ground of rape in Northern Ireland would not prevent a finding of an Article 3 violation. Furthermore, Zampas and Gher argue that, since Member States have a positive obligation to prevent ill-treatment, including rape, and to provide effective remedies, it can be asserted that this includes the right to access abortion in cases where pregnancy is the result of rape.[71]

CEDAW has taken a much clearer stance in finding that States should allow rape victims to access abortions. The landmark case *LC v Peru* concerned a thirteen-year-old girl who became pregnant from rape and attempted suicide by jumping from a building.[72] She was not allowed an abortion and was denied necessary surgery on the basis that it could harm the foetus, leaving her paralysed. CEDAW found that forcing the girl to continue her pregnancy "constituted cruel and inhuman treatment and therefore a violation of her right to physical, psychological and moral integrity".[73] The Committee recommended that Peru establish a mechanism for therapeutic abortion, and explicitly stated that abortion on the ground of rape should be decriminalised.[74] This judgment is significant in suggesting that prohibitions on abortion in the case of rape will always amount to inhuman and degrading treatment. There is thus significant jurisprudence under the ECHR and CEDAW to support the claim that Northern Ireland's prohibition on abortion on the ground of rape is a violation of Article 3 ECHR and Article 7 ICCPR.

C. Article 3 and FFA

Northern Ireland's prohibition on abortion on the ground of FFA also amounts to a violation of Article 3 ECHR and Article 7 ICCPR. There has only been one ECtHR case directly challenging the prohibition of abortion in cases of FFA, *D v Ireland,* but the Court avoided taking a position, finding the case to be inadmissible as the applicant had failed to exhaust domestic remedies.[75] However, in *R.R. v Poland*, the ECtHR found that the distress the applicant suffered as a result of being prevented from accessing information about the foetus' condition

[71] Zampas and Gher (n 58) 282.
[72] CEDAW, 'Views Communication No. 22/2009' (2011) CEDAW/C/50/D/22/2009.
[73] ibid [3.4].
[74] ibid [9(b)(i)] and [9(b)(iii)].
[75] *D v Ireland* App no 26499/02 (ECtHR, 27 June 2006).

was sufficient to reach the minimum threshold of severity under Article 3.[76] Again, foetal abnormality was already a lawful ground for abortion in Poland, but this case was the first in which the ECtHR considered Article 3 to be relevant in an abortion-related case. This, and the finding in *P and S*, shows a broadening in the approach of the Court and a diminishing reluctance to find restrictive abortion regimes in violation of the Convention. There is, therefore, the potential in a future case for the ECtHR to find that the prohibition on abortions on the ground of FFA violates Article 3.

This is even more likely in light of the recent HRC decisions in *Mellet v Ireland*[77] and *Whelan v Ireland*[78] that Ireland's restrictive abortion law, which forced the two applicants to travel to England for abortions in cases of FFA, violated the applicants' rights to freedom from inhuman and degrading treatment under Article 7 ICCPR. These cases are significant because of the similarities between the abortion regimes of Ireland and Northern Ireland.

Amanda Mellet was in her twenty-first week of pregnancy when she found out that the foetus she was carrying had a chromosomal abnormality that would result in its death *in utero* or shortly after birth. She travelled to Liverpool for an abortion at a high financial cost, did not seek aftercare when she returned to Dublin, was unable to access bereavement counselling services, and unexpectedly received the baby's ashes three weeks later. The Committee found that Mellet had been subjected to conditions of "intense physical and mental suffering" which was exacerbated by various factors including being forced "to choose between continuing her non-viable pregnancy or traveling to another country while carrying a dying foetus, at personal expense and separated from the support of her family, and to return while not fully recovered" and being subjected to "the shame and stigma associated with the criminalization of abortion of a fatally ill foetus".[79]

Rejecting Ireland's claim that there was no arbitrary interference with any right of the applicant, as there was no legal right to an abortion on the ground of FFA in Ireland, the Committee stated that the fact that particular conduct is legal under domestic law does not mean that it cannot infringe Article 7.[80] This right permits no limitations or justifications for its violation; accordingly, the Committee

[76] *R.R.* (n 62).

[77] UN Human Rights Committee, 'Views adopted by the Committee under article 5(4) of the Optional Protocol, concerning communication No. 2324/2013' (17 November 2016) CCPR/C/116/D/2324/2013.

[78] UN Human Rights Committee, 'Views adopted by the Committee under article 5(4) of the Optional Protocol, concerning communication No. 2425/2014' (12 June 2017) CCPR/C/119/D/2425/2014.

[79] *Mellet* (n 77) [7.4].

[80] ibid.

found that the facts of the case amounted to cruel, inhuman, or degrading treatment in violation of Article 7.[81] In light of this, a violation of Article 17 ICCPR (which mirrors Article 8 ECHR) was also found, as the State's interference in the applicant's decision as to how best to cope with her non-viable pregnancy was unreasonable and arbitrary.[82] It was therefore recommended that Ireland amend its law in compliance with the ICCPR to ensure effective, timely, and accessible procedures for abortion and, recognising that the existence of punitive provisions effectively amounted to the censorship of medical professionals, take measures to ensure that full information on safe abortion services can be provided without fearing criminal sanctions.[83]

Siobhán Whelan was in her twentieth week of pregnancy when she was informed that her foetus was affected by a rare brain malformation and a chromosomal condition and would die *in utero* or during or soon after birth. She travelled to Liverpool for an abortion at a total cost of €2,900. She was afraid to return to work and face questions, but was not entitled to paid maternity leave or grief counselling. She claimed that she had been subjected to cruel, inhuman, and degrading treatment in being denied the reproductive health care she needed, which forced her to continue to carry a dying foetus, compelled her to terminate her pregnancy abroad, and subjected her to intense stigma.[84] The Committee found that these facts established "a high level of mental anguish that was caused to the author by a combination of acts and omissions attributable to the State party" and violation of Article 7 was found.[85] The Committee again emphasised the absolute nature of Article 7, rejecting the State's explanations of moral and political considerations. A violation of Article 17 was also found along similar reasoning as *Mellet,* and the Committee stated that Ireland must take steps to prevent similar violations in future.

These decisions are of considerable relevance to Northern Ireland. In both cases, the Committee found that the prohibition on abortion in cases of FFA was unjustifiable, as it imposed a level of suffering amounting to inhuman and degrading treatment, exacerbated by forcing women to travel to a foreign country to obtain terminations with inadequate information and limited healthcare upon their return, and compounded by the existence of stigmatising criminal provisions. Northern Ireland similarly forces women to travel abroad to terminate non-viable foetuses and creates barriers to information due to the chilling effect of the criminal law, which imposes a sentence even higher than Ireland. While the

81 ibid [7.6].
82 ibid [7.8].
83 ibid [9].
84 *Whelan* (n 78) [3.1].
85 ibid [7.7].

UK has signed the ICCPR, it has not signed the first Optional Protocol to allow individual complaints to be brought before the HRC, so it is not possible for a case challenging Northern Ireland's abortion law to be brought under the ICCPR. However, *Mellet* and *Whelan* do still establish a human rights standard that the UK must comply with.

Furthermore, the two decisions could be influential in any future ECtHR case relating to the prohibition of abortion on the ground of FFA. The cases are relevant to the ECHR, as Article 7 ICCPR and Article 3 ECHR are equivalent to one another and Article 3 can be engaged in abortion-related cases, as demonstrated by *R.R.* and *P and S*. Article 3, as with Article 7, is an absolute right which permits no limitations, and therefore the margin of appreciation does not apply to this right. The Committee made clear in *Whelan* that arguments as to moral views (which the ECtHR accepted in *ABC* as engaging the margin of appreciation) could not be invoked as justification for the violation. As the threshold for inhuman and degrading treatment was met in *Mellet* and *Whelan*, it would be unjustifiable for the ECtHR to refrain from reaching the same conclusion in a future case relating to FFA, and it is therefore likely that the Court could find a violation of Article 3. It can thus be concluded that Northern Ireland must reform the law to allow for abortion in cases of FFA, as the UK is currently in breach of established human rights standards and risks being found in violation of Article 3 if such a case were brought before the ECtHR.

However, in the recent *NIHRC* case, only the a minority in the Supreme Court found an Article 3 violation.[86] The NIHRC initiated judicial review of Northern Ireland's abortion law in 2015, arguing that the failure to provide exceptions to the offence based on rape, incest, and fatal foetal abnormality violated Articles 3, 8, and 14.[87] The case was dismissed, as the majority took the view that the NIHRC did not have legal standing; had standing not been an issue, the Supreme Court would have found a breach of Article 8 regarding Northern Ireland's prohibition on abortion in cases of rape, incest, and FFA. Despite the *Mellet* and *Whelan* judgments recently preceding this case, only Lord Kerr and Lord Wilson found Northern Ireland's abortion law to violate Article 3. Lady Hale was of the view that Article 3 of the ECHR was more contextual than Article 7 of the ICCPR, and thus legislation could not be axiomatically regarded as a breach.[88]

However, Lord Kerr (with whom Lord Wilson agreed) felt that "a law requiring mothers to carry babies with fatal abnormalities to term or where their

[86] *The Northern Ireland Human Rights Commission's Application, Re Judicial Review* [2018] UKSC 27, [2019] 1 All ER 173.

[87] *The Northern Ireland Human Rights Commission's Application, Re Judicial Review* [2015] NIQB 96, [2015] 11 WLUK 798.

[88] *NIHRC* (n 86) [102]–[103].

pregnancy is the result of rape or incest, carries an inevitable risk that a number of them will have suffered inhuman or degrading treatment"[89] and, while not all women in such a situation would experience this degree of trauma, such a risk sufficiently amounts to an Article 3 violation where "proper safeguards to mitigate the risk of such trauma are not put in place."[90] Accordingly, Lord Kerr and Lord Wilson would have issued a declaration of incompatibility under section 4 of the Human Rights Act 1998 on this point. While the *Mellet* and *Whelan* cases were admittedly context-specific, the HRC identified a number of factors to support the finding of an Article 7 violation, including the lack of healthcare available in Ireland, forcing pregnant women to travel abroad, and the stigma of the criminal offence amounting to inhuman and degrading treatment. While, as Lord Kerr suggests, not every woman in this situation will experience such a level of trauma, it is evident that a significant number will experience treatment that could be categorised as inhuman and degrading; it is difficult to see how this would not be in contravention of Article 3.

There may be another opportunity, however, for the Supreme Court to issue a declaration of incompatibility on this point. At the end of January 2019, Sarah Ewart, a Northern Irish woman who travelled to England for an abortion in 2013 because the foetus she was carrying had anencephaly, brought a case before the Belfast High Court to challenge Northern Ireland's abortion law.[91] Since legal standing will not be an issue in this case, it is likely that the court will find Northern Ireland's abortion law to violate Article 8, and potentially Article 3, based on Ms Ewart's experience.

D. Decriminalisation

While it is important to recognise the need for abortion in cases of rape and FFA in Northern Ireland, only a small number of women require abortions on these grounds. Rachel Rebouché argues that the increasing support for a woman's right to an abortion, which is typically qualified to exceptional grounds, does not "capture the experiences of the majority of women or the reasons most women elect to end pregnancies."[92] Accordingly, there is a pressing need to fully decriminalise abortion in Northern Ireland, which can be advocated for based on

[89] ibid [223].
[90] ibid [235].
[91] BBC News, 'Abortion law: Sarah Ewart begins NI challenge' *BBC News* (London, 30 Jan 2019) <www.bbc.co.uk/news/uk-northern-ireland-47058629> accessed 14 Feb 2019.
[92] Rachel Rebouché, 'Abortion Rights as Human Rights' (2016) 25(6) *Social & Legal Studies* 765, 777.

human rights. As discussed above, the wide margin of appreciation afforded under Article 8 limits the scope of this right, so this argument will be based on Article 3.

The arguments for decriminalisation of abortion on the grounds of rape and FFA can be extended to the full decriminalisation of abortion, as the harms identified in the above cases are experienced by Northern Irish women seeking abortions on other grounds, whether they travel abroad or have illegal abortions. Fiona De Londras argues that Amanda Mellet's experience is applicable to women who want abortions for reasons other than FFA, and that the negative implications identified by the Committee "are all common experiences of abortion for women in Ireland' and all result in inhuman and degrading treatment."[93] In *Mellet* and *Whelan*, the finding of an Article 7 violation was based on the denial of healthcare and bereavement services in Ireland, the stigma caused by the criminal provisions, and the applicants being forced to travel abroad away from their support network at a great financial cost. These facts are relevant to most abortion-seeking women in Northern Ireland, and the harm suffered as a result amounts to inhuman and degrading treatment even where the foetus is viable.

The Committee emphasised that the Irish law forces women to "choose between continuing [their] non-viable pregnancy or traveling to another country",[94] but even where a foetus is viable, women are still compelled to continue an unwanted pregnancy or choose between travelling abroad or having an illegal abortion. The experiences in *Mellet* and *Whelan* are also common to women who do have illegal abortions, as the risk of prosecution acts as a barrier to information and can deter women from seeking aftercare and counselling services. Moreover, some women may not tell anyone in their support network about the abortion due to the stigma around abortion. Furthermore, the Committee found that many of the applicants' negative experiences could have been avoided if they had "not been prohibited from terminating [their] pregnancy in the familiar environment of [their] own country and under the care of the health professionals whom [they] knew and trusted",[95] which is true for all abortion-seeking women in Ireland and Northern Ireland.

The decriminalisation of abortion on the grounds of rape and FFA has been advocated for in several cases where violations of Article 7 were found. In *LC v Peru*, decriminalisation was explicitly recommended, and the HRC in *Mellet* and *Whelan* found that the criminalisation of abortion contributed to the suffering of the applicants by creating a barrier to information, healthcare, and support

93 Fiona De Londras, 'Fatal Foetal Abnormality, Irish Constitutional Law, and *Mellet v Ireland*' (2016) 24 *Med L Rev* 591, 606.

94 *Mellet* (n 77) [7.4].

95 ibid.

services. The HRC required that Ireland ensure "health-care providers are in a position to supply full information on safe abortion services without fearing being subjected to criminal sanctions".[96] It was recognised that the existence of a criminal offence of abortion exacerbated the harms suffered by the applicants, but this is not the case only in relation to rape and FFA; taking this further, Northern Ireland's criminal offence subjects women who require abortions on any ground to inhuman and degrading treatment, which makes full decriminalisation necessary. As the ECtHR cannot defer to the views of the State where a violation of Article 3 is found, there can be no justification for the breach. Thus, there is the potential for a case challenging Northern Ireland's abortion law under Article 3 to succeed, particularly in light of the progressive jurisprudence on abortion rights from both the ECtHR since *ABC* and recent decisions of the HRC. While the Supreme Court case failed on this point, the ECtHR could yet change this position.

While no individual complaint has yet resulted in this requirement, several UN bodies have recommended the decriminalisation of abortion, referencing Northern Ireland specifically. On the 23rd of February 2018, the CEDAW Committee released an in-depth report on Northern Ireland's abortion law, finding that the current position amounts to a "grave and systematic" violation of CEDAW.[97] The Committee acknowledged that "despite legal provision for abortion in very limited circumstances, de facto limitations render access to abortion virtually impossible",[98] and the criminalisation of abortion deprives Northern Irish women of any "real choice in influencing circumstances affecting their mental and physical health."[99] The "deplorable" options given to women in Northern Ireland force them to choose between "complying with discriminatory laws that unduly restrict abortion or [risking] prosecution and imprisonment",[100] a dilemma which is worsened by the lack of support structures for women who have had abortions and by the disregard for poor and socially vulnerable women and girls who face additional barriers. Following this finding, the reform of Northern

[96] ibid [9]; *Whelan* (n 78) [9].

[97] CEDAW, 'Report of the inquiry concerning the United Kingdom of Great Britain and Northern Ireland under Article 8 of the Optional Protocol to the Convention on the Elimination of All Forms of Discrimination against Women' (23 Feb 2018) CEDAW/C/OP.8/GBR/1, [1].

[98] ibid [20].

[99] ibid [42].

[100] ibid [81].

Ireland's abortion law is necessary to remedy the UK's ongoing violation of human rights.

E. Conclusion

Northern Ireland's abortion law must be reformed in order to comply with human rights standards. At the very least, the Polish cases and *ABC* require that abortion must be accessible on the grounds for which it is already lawful in Northern Ireland. Furthermore, based on the jurisprudence of the ECtHR, the CEDAW Committee, and the HRC, the prohibition on abortion on the grounds of rape and FFA amounts to a violation of Article 3 ECHR and Article 7 ICCPR. While reform is therefore required to guarantee access to abortion on these grounds, Article 3 and Article 7 can also be used to argue for the full decriminalisation of abortion, since the identified harms are suffered by all women requiring abortions in Northern Ireland and the right to freedom from inhuman and degrading treatment is an absolute one. Section 3 will consider the potential for reform in compliance with these rights, in order to guarantee access to abortion in cases of rape and FFA and the removal of the criminal offence.

III. Reforming Northern Ireland's Abortion Law

Recent statistics show overwhelming support for abortion in cases of rape and FFA in Northern Ireland, and while it is lower for more liberal grounds, there is still majority support for decriminalisation. It will now be considered how Northern Ireland's abortion can be reformed in accordance with the human rights standards established in Section II. As an absolute minimum, Northern Irish women must be able to access abortion services in cases of rape and FFA; however, legalisation on these cases alone would retain the restrictiveness of the law, meaning most abortion-seeking women would still be forced to travel abroad or have illegal abortions in violation of Article 3. Reform could therefore take the form of the extension of the Abortion Act or the provision of abortion on request in line with the European norm, but, as argued in Section II, the full decriminalisation of abortion would also be necessary under either alternative. Finally, law reform alone will not guarantee safe, legal abortions for all abortion-seeking Northern

Irish women; additional measures are necessary to make abortion accessible in practice.

A. Potential for an FFA ground

It was established in Section II that prohibitions on abortions in cases of rape and FFA amount to a breach of Article 3, so it is of critical importance that the law is reformed to allow abortions in these circumstances at the very least. It would be relatively easy to create an exception for FFA, which could be constructed broadly in terms of a *risk* that a foetal abnormality would be fatal. The Isle of Man's previous abortion legislation allowed for abortion if a medical practitioner was of the opinion that "there is a substantial risk that if the child were to be born at full term it would suffer from such physical or mental abnormalities as to be… unlikely to survive birth", as assessed by the good faith opinion of two medical practitioners.[101] Advancements in prenatal testing now allow doctors to assess with relative certainty the risk of FFA, so reliance on medical discretion would rule out the need for a concrete list of conditions or estimates of the length of time a baby would survive post-birth. Genetic testing is possible at an early stage of the pregnancy with very little risk to the pregnant woman or the foetus; non-invasive prenatal testing allows screening for foetal chromosomal abnormalities from the pregnant person's blood, which is possible from the tenth week of pregnancy. While the Committee on the Convention on the Rights of Persons with Disabilities (CRPD) has expressed that laws which explicitly allow for abortion on the grounds of impairment violate the CRPD,[102] an exception for *fatal* impairments, rather than all disabilities, focused on alleviating the suffering of pregnant women carrying fatally abnormal foetuses is less likely to be considered discriminatory. Thus, an exception for fatal foetal abnormalities, such as that previously used by the Isle of Man, would comply with human rights standards.

B. Rape exceptions are unworkable

However, rape exceptions are unworkable in practice. It would be impossible to require prosecution as evidence for allowing an abortion, because the delays that occur between reporting rape and the trial would exceed the length of the pregnancy. Furthermore, most cases do not even progress that far and the conviction rate for rape is extremely low; in 2017, there were 926 recorded rapes

[101] Termination of Pregnancy (Medical Defences) Act 1995, section 4(1)(b)(i).

[102] CRPD, 'Comments on the draft General Comment 36 of the Human Rights Committee on Article 6 of the International Covenant on Civil and Political Rights' (2017) [1].

in Northern Ireland, and only around fifty-nine (6.4%) resulted in charges.[103] It is well-established that rape trials subject complainants to "secondary victimisation" and most rapes are therefore not reported; one study estimated that in Britain, 80% of victims did not report their rape to the police and 29% did not even tell friends or family members.[104] The current Isle of Man legislation, while not requiring prosecution, does require the victim-survivor to present an affidavit taken under oath that the pregnancy was the result of rape; they must have made a report to the police "as soon was reasonable in all the circumstances"; and a medical practitioner must be satisfied that there is no medical evidence inconsistent with the allegation.[105] This still places an undesirable burden on rape victim-survivors to report rape immediately, and, as Fiona De Londras and Mairead Enright argue, a pregnant person should not have to satisfy the criminal burden of proof, which may cause further distress and degradation, in order to secure access to abortion.[106] Even where legislation does not require a police report, as in Germany where the law requires a "strong reason to support the assumption" that the pregnancy was caused by rape,[107] any burden of proof imposed still requires medical practitioners to believe the victim-survivor and could subject them to victim-blaming and scepticism as to the honesty of their complaint. Vicky Conway has argued that, as rape exceptions operate from the perspective of suspicion and cynicism, many victim-survivors may risk breaking the law by obtaining abortion pills, make the journey abroad, or carry the foetus to term against their wishes in order to avoid having to report their rape.[108] The onerous burden a rape ground would place women in the position where travelling abroad for an abortion, if they had the means to do so, is a more advisable option. A rape exception would thus be unworkable and would therefore fail to address the UK's human rights obligations.

C. Additional problems with exceptions – alternatives?

There are additional problems which make the introduction of new exceptions to the OAPA an inadequate option for reform. Simply creating exceptions

[103] Public Service of Northern Ireland, *Police Recorded Crime in Northern Ireland: Monthly Update to 31 December 2017* (2018).

[104] Nina Lakhani, 'Unreported rapes: the silent shame' *The Independent* (London, 12 March 2012) <www.independent.co.uk/news/uk/crime/unreported-rapes-the-silent-shame-7561636.html> accessed 5 February 2018.

[105] Termination of Pregnancy Act, section 5(2).

[106] Fiona De Londras and Mairead Enright, *Repealing the 8th* (Policy Press 2018) 98.

[107] German Criminal Code, section 218a(3).

[108] Vicky Conway, 'On The Difficulties of Rape Exceptions' (*Human Rights in Ireland*, 13 October 2014) <http://humanrights.ie/constitution-of-ireland/on-the-difficulties-of-rape-exceptions-repealthe8th/> accessed 26 February 2018.

for rape and FFA would retain the chilling effect of the law in contravention with Article 8. Given that doctors are reluctant to perform abortions even to save the life of a pregnant person, it is even more unlikely that they would be willing to use their discretion in these cases. In Poland, abortion is legal on the grounds of rape and foetal abnormality, but, as evidenced by the ECtHR cases, it is difficult for women to access abortion on these grounds in practice. Allowing abortions on the grounds of rape and FFA would not remedy the human rights violations established in Section II, as most Northern Irish women in these circumstances would continue to travel abroad or illegally terminate their pregnancies, either through being unable to access abortion services or wanting to avoid the trauma of satisfying a rape exception.

To guarantee access to abortion in such circumstances, it is necessary to provide for abortion on broader grounds, which could be done through the extension of the Abortion Act or by providing for abortion on request in line with the European standard. The extension of the Abortion Act would bring Northern Ireland in line with the rest of the UK, allowing women to access safe, legal abortions in Northern Ireland, relieving the criminal law of its chilling effect and no longer forcing women to travel abroad or risk prosecution and their health through illegal abortions. The social ground would encompass abortion on the ground of rape without imposing any evidential burden on the victim, and FFA would be covered by this ground up to twenty-four weeks and subsequently by the disability ground. However, the Abortion Act is a fifty-year-old statute and needs updating,[109] so a preferable alternative would be abortion on request. The European standard is twelve weeks, which Ireland has proposed, and after expiry of this period abortion is usually still permitted on limited grounds including foetal abnormality. Both options would satisfy the UK's human rights obligations under Article 3.

D. Decriminalising abortion

As argued in Section II, the full decriminalisation of abortion is also necessary. It is important to consider what the purpose of a law criminalising abortion is; Marge Berer argues that criminal provisions "make sense only for punitive and deterrent purposes, or to protect fetal life over that of women's lives",[110] and Sally Sheldon has noted that the two commonly alleged purposes of the Abortion Act are to prevent the destruction of foetal life and to prevent harm

[109] See Sally Sheldon, *Beyond Control: Medical Power and Abortion Law* (Pluto Press 1997).

[110] Marge Berer, 'Abortion Law and Policy Around the World: In Search of Decriminalisation' (2017) 19(1) *Health and HRJ* 13, 14.

to women.[111] Accordingly, the maximum life sentence imposed by the OAPA is intended to have a deterrent effect in order to protect foetal life. This is premised on particular moral beliefs about foetal life, which is not only an inappropriate basis for a severe criminal offence, but also ignores modern views towards abortion. As established in Section I, the majority of the Northern Irish population support access to abortion in at least some cases. Sally Sheldon argues that moral views alone are "insufficient to ground a criminal prohibition",[112] and the claimed deterrent function of the offence cannot be used to bolster these moral convictions as a justification, since the offence has proven to be an ineffective deterrent. As outlined above, although the criminal law does have a chilling effect on doctors, thousands of women continue to have abortions, whether illegally or by travelling abroad. Punitive provisions clearly do not achieve the aim of protecting foetal life, and thus cannot be justified on this basis.

Preventing harm to women was part of the rationale behind the provisions of the 1861 Act, since at the time abortion was dangerous and carried a significant risk of death or injury. However, medicine has advanced to the point where early medical abortion is now the norm and is generally safer than pregnancy and childbirth. Furthermore, as highlighted in the first two sections, Northern Ireland's criminal provisions result in harsh prosecutions, the imposition of financial and emotional burdens on women that travel abroad, health risks due to illegal abortions, and human rights violations. Berer points out the irony of arguing that restrictive laws protect women when it is those laws that are "responsible for the deaths and millions of injuries to women who cannot afford to pay for a safe illegal abortion."[113] Even where abortion has been legalised on broader grounds, the impacts of the criminal law are still felt. While the Abortion Act mitigated most of the force of the 1861 Act, prosecutions for illegal abortions in England still occur; in 2013, one woman was sentenced to three-and-a-half-years in prison (reduced from an initial sentence of eight years) for terminating her pregnancy at thirty-nine weeks,[114] and, in 2015, another woman received a two-and-a-half-year sentence for terminating her thirty-four-week pregnancy.[115] Prosecutions in England are admittedly rare, and there is no recent evidence of doctors being prosecuted for performing abortions in Northern Ireland, but the existence of a criminal offence still creates that risk. Furthermore, the criminalisation of abortion

[111] Sally Sheldon, 'The Decriminalisation of Abortion: An Argument for Modernisation' (2015) 36(2) *Oxford Journal of Legal Studies* 334, 347.

[112] ibid 351.

[113] Berer (n 110) 14.

[114] *R v Catt* [2013] EWCA Crim 1187, [2013] 6 WLUK 263.

[115] BBC News, 'Shildon woman jailed over poison termination' *BBC News* (London, 17 December 2015) <www.bbc.co.uk/news/uk-england-tyne-35121524> accessed 26 February 2018.

stigmatises women, and women that choose to have abortions are often labelled "murderers" by anti-abortion activists. The OAPA fails at its aim of protecting foetal life while causing harm to women; therefore legal reform must be aimed at abolishing the outdated criminal offence of abortion.

E. Additional requirements to abortion access

Decriminalisation alone will not automatically guarantee all women a right to an abortion—there is a need for further measures to ensure that abortion is accessible in practice. Canada and four Australian States (the Australian Capital Territory, Victoria, Tasmania, and the Northern Territory) are the only places worldwide to have decriminalised abortion, but there remain issues in accessing abortion. Abortion was decriminalised in Tasmania in 2013, but since then access to abortion services has been significantly reduced due to the closing of two clinics on the island and public hospitals remaining reluctant to carry out abortions.[116] In Canada, abortion was decriminalised in 1988, when the Supreme Court struck down the country's restrictive abortion legislation.[117] There is no federal law regulating abortion, and it is instead considered to be a "medically necessary" service.[118] However, one study showed that women living in rural, Northern, and coastal communities had to travel a considerable distance for abortions, sometimes more than 100km.[119] The abortion pill was not made available in Canada until 2017, and access to it remains limited, with less than half of all Canadian provinces supplying the drug.[120] Thus, Barbara Baird argues that "while decriminalization may be a precondition for the improvement of access to abortion services, it is only when public health departments take responsibility that equitable access will be delivered."[121]

Therefore, alongside legal reform, action must be taken by medical and social institutions to ensure access to abortion. Katherine Side argues that efforts are required to ensure that there are an adequate number of doctors willing to refer women and perform abortions and that abortion is an economically

[116] Barbara Baird, 'Decriminalization and Women's Access to Abortion in Australia' (2017) 19(1) *Health and HRJ* 197, 203–204.

[117] *R v Morgentaler* [1988] 1 SCR 30.

[118] Christabelle Sethna and Marion Doull, 'Spatial disparities and travel to freestanding abortion clinics in Canada' (2013) 38 *Women's Studies International Forum* 52.

[119] ibid 59–60.

[120] Ashley Csandy, 'Abortion pill available in less than half of all Canadian provinces three months after rollout' *National Post* (Toronto, 30 March 2017) <http://nationalpost.com/news/canada/abortion-pill-available-in-less-than-half-of-all-canadian-provinces-three-months-after-rollout> accessed 26 February 2018.

[121] Baird (n 116) 198.

accessible service for all women.[122] This would require the sufficient training of doctors to perform abortions, funding for the provision of NHS abortion services, access to contraception, and access to medical abortions, including the option for the abortion pill to be taken at home, so that women are not required to travel to and from the clinic over two days to take the drugs. Such a move was approved in Scotland last year,[123] despite an unsuccessful legal challenge from the SPUC,[124] and in England[125] and Wales[126] more recently. Furthermore, the recognition of the right to conscientious objection should not impede access to abortion services. While it is important to respect the religious or moral beliefs of doctors, this cannot come at the cost of unequal abortion provision. In the rest of the UK, doctors who conscientiously object are not required to disclose that this is the reason for denying the woman's request; nor are they obligated to refer the woman to another doctor who would not object on this ground. Emily Jackson expressed concern that some women could mistake this for an indication that they are ineligible for an abortion, particularly those with limited knowledge of the abortion legislation or the lack of confidence to seek a second opinion; even if a woman did then seek a second opinion, this would invariably cause a delay.[127] De Londras and Enright argue that a woman seeking an abortion should be entitled to know whether a doctor would object on grounds of conscience, to save them time and any potential distress; this could be achieved by conducting a census and registering the doctors that would conscientiously object.[128] This would provide an appropriate balance between the rights of doctors and the rights of women seeking abortions.

Adequate access to abortion also requires universal access to information and counselling services, and Northern Ireland would also require more widespread access to prenatal screening. Women should be able to access services free of shame and stigma; decriminalisation would be a step towards this, but additional "buffer zones" could be imposed around abortion clinics to prevent anti-abortion activists from protesting within a certain radius. The CEDAW Committee identified that the social context of abortion in Northern Ireland, which is dominated by the

[122] Side (n 4) 103.

[123] Libby Brooks, 'Women in Scotland will be allowed to take abortion pill at home' *The Guardian* (London, 26 October 2017) <www.theguardian.com/world/2017/oct/26/women-scotland-allowed-take-abortion-pill-at-home> accessed 26 February 2018.

[124] *SPUC Scotland, Re Judicial Review* [2018] CSOH 85, [2018] 8 WLUK 140.

[125] Haroon Siddique, 'Use of second abortion pill at home to be allowed in England' *The Guardian* (London, 25 August 2018) <www.theguardian.com/world/2018/aug/25/use-of-second-abortion-pill-at-home-to-be-allowed-in-england> accessed 25 August 2018.

[126] BBC News, 'Women allowed to take abortion pill at home in Wales' *BBC News* (London, 29 June 2018) <www.bbc.co.uk/news/uk-wales-44643459> accessed 25 August 2018.

[127] Emily Jackson, *Regulating Reproduction* (Hart 2001) 86.

[128] De Londras and Enright (n 106) 80–81.

anti-abortion rhetoric of politicians and religious figures, coupled with the lack of family planning places women in "double jeopardy"[129] which is worsened by the presence of anti-abortion protesters harassing women seeking abortions.[130] The Committee recommended urgently that sections 58 and 59 of the OAPA be repealed and legislation be enacted to allow abortions on health grounds without requiring long-term or permanent effects on the woman's health, as well as abortions in cases of rape and severe foetal impairment.[131] The provision of adequate sexual and reproductive health services, including protection against anti-abortion protesters, is also recommended.[132] The human rights of abortion-seeking women in Northern Ireland will therefore only be realised fully where abortion is decriminalised and additional requirements are put in place to ensure that women are able to access safe, legal abortion services in practice, free from any financial or emotional burden or stigma.

F. Conclusion

While it is necessary to provide for abortions in cases of rape and FFA, reformulating the law to include these grounds alone would not fully address the human rights issues within Northern Ireland's abortion law. That approach would retain the chilling effect of the criminal offence, and the alternative of a rape exception would be unworkable as it would place an onerous burden on rape victims, so that travelling abroad would continue to be the best option. A more adequate alternative would be to extend the Abortion Act or provide for abortion on request for up to twelve weeks of pregnancy. The full decriminalisation of abortion is also necessary, as the OAPA fails to achieve its purported aims, creates a chilling effect on doctors, harms women, and perpetuates abortion stigma. Alongside decriminalisation, extra-legal measures would also be necessary to ensure that access to abortion is unimpeded.

Conclusion

Northern Ireland's abortion law is exceptionally restrictive compared with the rest of Europe, yet Northern Irish politicians have remained unwilling to update the 150-year-old legislation guiding the country's abortion regime due to the perceived dominance of anti-abortion views. This has also given Westminster an excuse for inaction, despite overwhelming evidence that the majority of Northern

[129] CEDAW (n 97) [51].
[130] ibid [19]–[20].
[131] ibid [85].
[132] ibid [86].

Ireland is actually in support of liberalised abortion laws. The chilling effect of the criminal offence means that doctors are unwilling to perform abortions even on lawful grounds, forcing women to travel abroad or have illegal abortions in virtually all circumstances. It has been argued that Northern Ireland's abortion law violates Article 8 due to the chilling effect of the OAPA and Article 3 due to the prohibition of abortion on the grounds of rape and FFA and the criminalisation of abortion. Reform must ensure access to abortion in cases of rape and FFA, but the full decriminalisation of abortion is also necessary and non-legal measures such as additional funding, the training of doctors, and buffer zones are essential in order to fully realise the human rights of abortion-seeking women.

One currently significant barrier to abortion law reform is the fact that the Northern Irish Assembly has been without an executive since January 2017, and negotiations to restore power-sharing between the two main political parties have failed.[133] The Northern Ireland Health Department completed a report on abortion in cases of FFA in 2016, but due to the lack of an executive to approve the document, the report was only released in April 2018. The report confirmed the need for a change in the law and recognised the serious adverse effects the current position could have on a woman's wellbeing, but with no executive, the Northern Irish Assembly cannot act upon this.[134] Considering the reluctance of previous UK governments to extend the Abortion Act in the past and Theresa May's recent deal with the Democratic Unionist Party,[135] which takes a strong anti-abortion position, it is unlikely that the current government would be willing to liberalise Northern Ireland's abortion law, even if power-sharing is restored to Westminster. However, the CEDAW Committee has stated that devolution cannot be used as a justification for the failure of the UK government to protect human rights;[136] Westminster has a responsibility to address the violations resulting from Northern Ireland's abortion law, regardless of the Northern Irish political situation. Efforts have been made to introduce equal marriage to Northern Ireland in the absence of the Northern Irish executive,[137] and, following the Irish referendum, Stella Creasey tabled an emergency debate in Parliament to raise the issue of abortion in Northern Ireland

[133] Henry McDonald, 'May fails to reach deal to restore Northern Ireland power-sharing' *The Guardian* (London, 12 February 2018) < www.theguardian.com/uk-news/2018/feb/12/theresa-may-arrives-in-belfast-for-stormont-assembly-talks> accessed 27 February 2018.

[134] Department of Health NI, *Healthcare and the Law on Termination of Pregnancy for Fatal Fetal Abnormality* (2016).

[135] BBC News, 'Conservatives agree pact with DUP to support May government' *BBC News* (London, 26 June 2017) <www.bbc.co.uk/news/uk-politics-40403434> accessed 27 February 2018.

[136] CEDAW (n 97) [52]-[53].

[137] BBC News, 'Same-sex marriage legislation for NI' *BBC News* (London, 28 March 2018) <www.bbc.co.uk/news/uk-northern-ireland-43570541> accessed 31 March 2018.

and advocate for decriminalisation across the UK.[138] The Women and Equalities Committee also launched an inquiry into abortion law in Northern Ireland,[139] and while this may not result in legal change, the increasing focus on this issue is making it harder for the government to continue avoiding the fact that the current legal position is indefensible, and reform is long overdue.

[138] Heather Stewart, 'MPs call on May to decriminalise abortion in Northern Ireland' *The Guardian* (London, 3 June 2018) <www.theguardian.com/uk-news/2018/jun/03/mps-call-on-may-to-decriminalise-abortion-in-northern-ireland> accessed 9 July 2018.

[139] The Women and Equalities Committee, 'Inquiry: Abortion Law in Northern Ireland' (*www.Parliament.uk*, 20 September 2018) <www.parliament.uk/business/committees/committees-a-z/commons-select/women-and-equalities-committee/news-parliament-2017/abortion-law-northern-ireland-launch-17-19/> accessed 4 October 2018.

Legal Personality of Ganga and Ecocentrism: A Critical Review

*Palash Srivastav**

I. Introduction

The river Ganga drains a vast expanse of Indian territory, originating from the Gangotri Glacier in the Himalayas and flowing through five Indian states. Running for a total distance of 2,525 kilometres, Ganga carries the distinction of being the longest river in India. Due to its religious and commercial importance, moreover, Ganga has been an important subject of law and policy. Ganga is therefore of considerable political importance to the current regime. The Prime Minister of India, Narendra Modi, contested the 2014 Lok Sabha election from Varanasi, a city of religious importance situated on the banks of Ganga, by claiming that he had come there on "Maa Ganga's call" and accordingly promised to have Ganga cleaned up within four years of being elected.[1] Once his government was sworn in, one of its first executive decisions was to rename the Ministry of Water Resources as the Ministry of Water Resources, River Development and Ganga Rejuvenation, asserting that rejuvenation of Ganga was a top priority of the government.[2] Ganga

* Fourth year student at the National Law School of Indian University. I would like to thank Professor Abhayraj Naik for carefully reading and suggesting changes to the initial drafts of this paper, and for guiding my research on ecocentrism in the Indian context.

1 Rajiv Srivastava, 'Modi files nomination, says 'Ma Ganga has called me to Varanasi' *The Times of India* (Mumbai, 24 April 2014) <https://timesofindia.indiatimes.com/news/Modi-files-nomination-says-Ma-Ganga-has-called-me-to-Varanasi/articleshow/34161157.cms> accessed 20 March 2019.

2 Sanjib Kr Baruah, 'Modi govt to change names of several key ministries' *Hindustan Times* (New Delhi, 7 August 2014) <https://www.hindustantimes.com/india/modi-govt-to-change-names-of-several-key-ministries/story-cfmUpxgiG9nETRWOY2igGP.html> accessed 20 March 2019.

has also been declared to be the National River of India,[3] and a draft National River Ganga Bill 2018 (Ganga Bill) has been floated by the Ministry to introduce strong protective measures for the river, including an armed investigative force and a statutory authority to carry out environmental impact assessments for proposed projects.[4]

For years, however, Ganga's condition has been constantly deteriorating. The highly dammed and polluted river is slipping into a state of advanced degradation, despite the billions of rupees being spent by the Indian government toward its rejuvenation.[5] Moreover, the state maintains a repressive or apathetic approach to activists fighting for Ganga's rights. With the recent demise of professor and environmental activist Guru das Agarwal, who died while fasting to draw the government's attention to the plight of Ganga, burning questions pertaining to the river have yet again come to the fore, such as how it can best be protected and what can be done to halt pollution. Legal solutions to these questions are imperative to ensure continued sustenance and ecological flow of the river.

In March 2017, two successive judgments of the Uttarakhand High Court (High Court) conferred legal personhood on Ganga, another major river Narmada, and their ecosystems.[6] To have legal personhood is to be recognised by law as a holder of rights and duties. Legal personality for natural entities and ecosystems gives them enforceable rights in law, and the ability to have standing in their own name or through proxies to enforce their rights in court. To the extent that recognition of claims is a key facet of justice, personality in law of a natural entity or of an ecosystem is a crucial step for the legitimation of its claims and needs.

This paper focuses on the High Court's rulings and observations pertaining to Ganga. The first section discusses the High Court's ruling in each case, and the material bases on which the High Court granted legal personality to Ganga.

[3] Know India, 'National River' <https://archive.india.gov.in/knowindia/national_symbols.php?id=7> accessed 28 November 2018; Ministry of Water Resources, 'Draft National River Ganga Bill, 2018' (*Press Information Bureau*, 4 February 2019) <http://pib.nic.in/newsite/PrintRelease.aspx?relid=188011> accessed 31 March 2019 (Draft Ganga Bill).

[4] Jayashree Nandi, 'River Ganga may get armed guards with powers to arrest polluters' *Hindustan Times* (New Delhi, 1 November 2018) <https://www.hindustantimes.com/india-news/draft-bill-on-ganga-proposes-armed-corps-for-protection-of-river/story-aaoha7O5uNSZmrIzbzYyzJ.html> accessed 28 December 2018.

[5] Susil Sahoo, 'Project-wise Amount Sanctioned and Amount Spent under National Mission for Clean Ganga as on 30-06-2017' (*Government of India*, 21 August 2018) <https://community.data.gov.in/project-wise-amount-sanctioned-and-amount-spent-under-national-mission-for-clean-ganga-as-on-30-06-2017/> accessed 20 March 2019.

[6] *Salim v State of Uttarakhand* WP (PIL) No. 126/2014 (Uttarakhand High Court); and *Miglani v State of Uttarakhand* WP (PIL) 140/2015 (Uttarakhand High Court).

This is followed in the second section by a thorough examination of the reasoning employed in the judgments to argue that no effective personhood has been conferred on Ganga. The final section explores ways in which the shortcomings of the judgments may be rectified to confer meaningful legal personality on Ganga. To that end, this paper proposes to read the High Court judgments in light of the Supreme Court judgments that have recognised ecocentrism as a principle of Indian law, and attempts to locate the judgments within the broader context of a discursive change happening in environmental jurisprudence. Such a reading would allow for some of Ganga's rights, like the right to life, to be recognised as being enforceable in a court of law. Christopher Stone's framework is employed in this paper to unpack the implications of Ganga's legal identity. Stone's seminal work arguing for legal recognition of natural entities remains relevant to this day and has been used to examine the substance of legal personality of natural entities in other jurisdictions.[7]

The Ganga Bill, the text of which is still not publicly available, must take these developments on the rights of Ganga into account and incorporate the ecocentric gains that have been made in the jurisprudence on rights of natural entities. At the time of writing, what little is known of the Ganga Bill from the written reply to Parliament by the Minister for Water Resources does not inspire confidence that the Bill will be a pioneer in recognising rights of natural entities from an ecocentric perspective, despite the rise of the ecocentrism as a legal discourse in India.[8] This paper hopes to draw together jurisprudential developments and thereby provide meaningful guidance to legislative efforts toward effective legal personhood for Ganga.

II. Overview of the Uttarakhand High Court Judgments

The two judgments of the High Court in *Salim v State of Uttarakhand* and *Miglani v State of Uttarakhand*, delivered respectively on 20 and 30 March 2017, granted legal personality to Ganga.[9] While in the case of *Salim* a division bench of the High Court had conferred the status of a legal person on Ganga river,[10] in the case of *Miglani* the same division bench conferred legal personality not just on

[7] See for example James DK Morris and Jacinta Ruru, 'Giving Voice to Rivers – Legal Personality as a Vehicle for Recognising Indigenous Peoples' Relationships to Water?' (2010) 14(2) AILR 43.

[8] Draft Ganga Bill (n 4).

[9] *Salim* (n 6); and *Miglani* (n 6).

[10] *Salim* (n 6) [19].

the river but on its entire ecosystem, including mountains, glaciers, and streams.[11] Even though the former judgment has been stayed by an order of the Supreme Court of India,[12] the latter remains binding law in the State of Uttarakhand.[13] While the judgments are certainly laudable in their ambition, the legal sources cited by the Judges to confer legal personhood on Ganga leave much to be desired in terms of fashioning substantive legal personhood for the river.

In *Salim*, the High Court invoked its parens patriae jurisdiction to make Ganga a legal person.[14] Parens patriae is a doctrine, meaning "parent of the nation", usually invoked by a court on application of the executive government to protect persons whom it thinks cannot advance their own interests, such as minors and legally incompetent persons. The High Court cited Ganga's religious importance as a major reason for conferring personality on it,[15] drawing on a long history of courts doing so in respect of Hindu deities.[16] The High Court further observed that the "socio-political-scientific" development of the country was a key driver in extending legal personality to Ganga.[17] Finally, the High Court drew upon the constitutional duties of the state and every citizen 'to protect and improve the environment' to justify the extension of personhood to the river.[18] The High Court accordingly appointed the Director of Namami Gange,[19] the Chief Secretary of the State of Uttarakhand, and the Advocate General of the State of Uttarakhand as "*persons in loco parentis*", or guardians of Ganga, making them the "human face" of Ganga to protect, conserve and preserve the river, and to uphold its status and to promote its health and well-being.[20]

In *Miglani*, the same division of the High Court heard another petition to declare the "Himalayas, Glaciers, Streams, Water Bodies etc." of Ganga, along with the Gangotri Glacier that feeds Ganga, as legal persons.[21] In its judgment,

[11] *Miglani* (n 6) [62].

[12] *State of Uttarakhand v Salim* 2017 SCC Online SC 903.

[13] The Uttarakhand High Court has power of supervision over all subordinate courts located in the territory over which it has jurisdiction, giving it the power to enforce its ruling across this territory: Constitution of India 1950, art 227.

[14] *Salim* (n 6) [11].

[15] ibid [17].

[16] ibid [12]–[15], citing *Naskar v Commissioner of Income Tax, Calcutta* 1969 (1) SCC 555 (Supreme Court of India), *Ram Jankijee Deities v State of Bihar* 1999 (5) SCC 50 (Supreme Court of India), *Shiromani Gurudwara Prabandhak Committee, Amritsar v Dass* AIR 2000 SC 1421 (Supreme Court of India), and *Behari Ji v Dass* AIR 1972 Allahabad 287 (Allahabad High Court).

[17] *Salim* (n 6) [16].

[18] ibid [18], citing Constitution of India (n 13) arts 48A and 51A(g).

[19] Namami Gange, or the National Mission for Clean Ganga, is a flagship programme of the central government for the abatement of pollution and rejuvenation of Ganga.

[20] *Salim* (n 6) [19] (emphasis in original).

[21] *Miglani* (n 6).

the High Court reaffirmed the status of Ganga as a legal entity. The reasoning employed by the High Court shares broadly common grounds with *Salim*. First, the High Court saw such a conferral of legal personality on Ganga as attuned with the constitutional, legal and moral duty of citizens to preserve and protect nature.[22] Secondly, the High Court asserted that under the "New Environment Justice Jurisprudence" and the principles of parens patriae, it is the courts' duty to protect nature.[23] Thirdly, the High Court saw bestowal of legal personality on Ganga as symbolic recognition of the "Constitutional legal rights" of "Mother Earth".[24] Fourthly, intergenerational equity weighed with the High Court as a factor necessitating grant of legal personality on Ganga for its continued existence.[25] Finally, the High Court again cited "socio-political-scientific" development as the backdrop to evolving conceptions of legal personality.[26] Building upon *Salim*, the judgment in *Miglani* appointed additional authorities as guardians of the river and its ecosystem.[27]

While both judgments declared Ganga to be a legal person for broadly the same reasons, the High Court did not take its declarations to their logical conclusion by fleshing out the content of Ganga's legal personality. The next section adopts the perspective of environmental justice to highlight pitfalls in the reasoning of the High Court and attempts to show how no meaningful legal personhood has yet been conferred on Ganga.

III. Critique of the Judgments from an Environmental Justice Perspective

An examination of the reasoning of the judgments shows that the legal personhood of Ganga, as crafted by the High Court, is fraught with problems. This paper adopts an environmental justice approach to critique the judgments in order to place the affected ecosystems and dependent people and animals at the centre of its analysis.[28] The purpose of this section is to bring environmental

[22] ibid [39].
[23] ibid [39].
[24] ibid [55].
[25] ibid [59].
[26] ibid [60].
[27] Additional appointees included Director of National Mission for Clean Ganga, the Legal Advisor to Namami Gange Project, the Director (Academics) of the Chandigarh Judicial Academy, and noted environmental lawyer Senior Advocate MC Mehta were appointed as guardians of the Ganga: ibid [62.3].
[28] David Harvey, *Justice, Nature and the Geography of Difference* (Wiley-Blackwell, 1997) 385–387.

perspectives to bear on the reasoning adopted by the High Court to examine the content of Ganga's newfound legal personhood.

The judgments are problematic for many reasons, but five stand out. First, the High Court invoked the doctrine of parens patriae erroneously to confer personality on Ganga even when it does not have the unilateral authority to invoke it. Secondly, Ganga has effectively been cast by the High Court as a perpetual minor who lacks full-fledged legal personhood, as opposed to the developed personhood of other legal entities. Thirdly, although it is apparent from the language of the judgments that the High Court was aware of international developments in the rights-of-nature movement, it did not meaningfully engage with these developments and thus deprived itself of the opportunity to flesh out a legally coherent conception of personhood for Ganga. Fourthly, the guardians appointed by the High Court lacked independence, accountability and substantive powers to ensure Ganga's protection. Finally, the High Court judgments fall foul of secular values enshrined in the Indian Constitution. This section elaborates on each of these heads of criticism.

A. *Parens patriae* has been wrongly invoked by the High Court

As discussed, the High Court invoked the doctrine of parens patriae to grant personhood to Ganga. While the doctrine was mentioned only in passing in *Salim*, the Court in *Miglani* dealt with it extensively, quoting vastly from judgments and academic texts in the United States. Even in the cited judgments, however, the doctrine is not invoked by courts to confer authority to act on behalf of either natural or legal persons without an application by the state. The courts, while discharging their judicial function, are not considered part of the "State" under the Indian Constitution.[29] Indeed, parens patriae is exercised conventionally on the application of the executive branch. While the doctrine has been relied upon by the executive in cases involving environmental questions,[30] its unilateral invocation

[29] Constitution of India (n 13) art 12.

[30] Outside of litigation, the doctrine of parens patriae was most famously invoked by the central government in the aftermath of the Bhopal Gas disaster as a justification for passing the Bhopal Gas Leak Disaster (Processing of Claims) Act 1985 (India). The effect of this Act was that the executive became the only authority that could legitimately seek compensation on behalf of the victims, depriving them of the right to litigate on their own behalf.

by a court in conferring authority over and responsibility for a natural entity on government personnel departs starkly from precedent.

B. Ganga has been cast as a perpetual minor without sufficient autonomy

The kind of guardianship that has been fashioned for Ganga by the High Court in exercise of its parens patriae jurisdiction casts Ganga as a perpetual minor with a curtailed personhood as opposed to personhood of other legal entities such as companies.[31] In a company, the board of directors is accountable to the shareholders who may remove directors on their own volition.[32] No parallel power for removal of Ganga's guardians has been reposed in any person or group; the guardians cannot be removed at all, save by the authority of the High Court or the Supreme Court of India. The judgments are completely silent on the accountability of the guardians, and what action would make them liable for being stripped of their designation. For removal of a guardian under Guardians and Wards Act 1890 (GAWA), a person must first apply to the court and show that the conduct of the guardian falls under one of the ten heads which makes the guardian liable for removal.[33] While the process for removal of a director tends to remain an internal matter of the company which does not involve the court, the necessary involvement of courts in removal of guardians has the potential for tiresome litigation. The High Court's reliance on the *parens patriae* doctrine has fashioned Ganga as a perpetual minor, bereft of the rights, protections and functional autonomy that a full-fledged legal person, such as that which a company enjoys.

C. Absence of sustained engagement with major developments in the rights of nature movement

While the High Court in *Miglani* cites the "New Environment Justice Jurisprudence" as a major reason for conferring legal personality on river Ganga,[34] it does not shed any light on what it means by this phrase. Both of these judgments are delivered close to the time when Te Awa Tupua (Whanganui River Claims Settlement) Act 2017 (NZ Act) was enacted in New Zealand.[35] The New Zealand

[31] Erin L O'Donnell and Julia Talbot-Jones, 'Legal rights for rivers: what does this actually mean?' (2017) 32(6) Australian Environment Review 159, 161.

[32] Companies Act 2013 (India), s 169.

[33] Guardians and Wards Act 1890 (India), s 39.

[34] *Miglani* (n 6) [39].

[35] Te Awa Tupua (Whanganui River Claims Settlement) Act 2017 (NZ Act) received royal assent on the same date as the High Court delivered its *Salim* judgment.

Parliament, through this Act, conferred legal personhood on the entirety of the Whanganui River and appointed a representative guardian body — Te Pou Tupua — for the protection of the river's rights. The guardian body consists of two representatives, one appointed by the Crown and the other appointed by the Whanganui Iwi (tribe) and has strong powers to pursue its protective functions.[36]

With a dedicated fund of $30 million, an advisory group and a strategy group for the implementation of the catchment management strategy,[37] the NZ Act ensures independence and institutional support for the guardian body to act in the best interest of the Whanganui River. This form of independence is missing from the framework for personhood created by the High Court to protect Ganga. While the High Court in *Miglani* borrowed language from the NZ Act,[38] it does not seem to take inspiration from the spirit of autonomy it embodies. The judgments also do not mention other international legal developments that have protected nature and natural entities by conferring legal personality on them, such as the changes made to the Ecuadorean and Bolivian laws to recognise rights of nature.[39] The High Court recognised the need to confer constitutional rights on "Mother Earth"[40] much along the lines of the Bolivian "Law of Mother Earth",[41] and the Universal Declaration of Rights of Mother Earth adopted by the World People's Conference on Climate Change and the Rights of Mother Earth in Bolivia in 2010,[42] making it likely that the High Court was aware of these developments.

While the language and the mechanisms created under these laws cannot be identically transplanted to India, given the Ganga's uniqueness, a sustained discussion of these developments abroad would have certainly contributed to the substance of legal personality in terms of its implications, the nature of rights that can be held by Ganga and mechanisms for their enforcement. The Ecuadorean and Bolivian laws could have provided rich source material for the development

[36] Erin L O'Donnell and Julia Talbot-Jones, 'Creating legal rights for rivers: lessons from Australia, New Zealand, and India' (2018) 23(1) Ecology and Society 7, 10.

[37] O'Donnell and Talbot-Jones, 'Legal rights for rivers' (n 41) 160. See generally NZ Act (n 45) ss 18–19.

[38] The formulae "human face", for instance, is lifted verbatim from s 18 of the NZ Act (n 45).

[39] Ecuador recognised the rights of nature in its Constitution through a 2008 amendment and by 2010 legislation in Bolivia: Lidia Cano Pecharroman, 'Rights of Nature: Rivers That Can Stand in Court' (2018) 7(1) Resources 13, 16–17.

[40] *Miglani* (n 6) [55].

[41] Ley de Derechos de la Madre Tierra (071) (Bolivia).

[42] Universal Declaration of Rights of Mother Earth (22 April 2010) <http://therightsofnature.org/universal-declaration/> accessed 8 March 2019.

of Ganga's legal personality.[43] By not accounting for the juridical machinery underpinning the borrowed phraseology, the High Court deprived itself of the opportunity to elaborate Ganga's legal personhood in a more effective manner.

D. Lack of independence and non-suitability of Ganga's guardians

The High Court's designation of seven persons, of which five were government officials,[44] as the guardians of Ganga should be looked at with circumspection. The state is often complicit in defiling Ganga, as demonstrated by the number of dams approved by the central government and the unanswered letters from Professor Agarwal to Prime Minister Modi.[45] In a comparable context, GAWA declares that a fiduciary relationship exists between the guardian and the ward,[46] and the existence of any interest that is adverse to their duties as the guardian makes them liable for removal.[47] Ganga's bureaucrat guardians, such as the Chief Secretary and the Advocate General of the State of Uttarakhand, may find themselves in positions of conflict with their fiduciary duty to Ganga due to the nature of their offices. These conflicts of interest are likely to occur when, for example, the Chief Secretary must ensure the implementation of a decision of the Uttarakhand government pertaining to Ganga which might not be in the river's best interest, or when the Advocate General must appear in court to defend the government's actions pertaining to Ganga, whatever their nature may be.

[43] The Ecuadorean Constitution specifically recognises nature as a holder of rights that are designed for it (art 10). It also empowers "all persons, communities, peoples and nations" to move governmental authorities whenever Constitutional rights of nature are violated (art 71). The Bolivian law seeks to recognise the earth and its living systems as living entities entitled to equal protection as humans under the law (art 5). The Bolivian law also recognises the right to life of earth and of its living systems, and fastens certain duties on the "Plurinational State" (art 8) as well as "natural persons and public or private legal entities" (art 9) for the protection of Mother Earth and of its components. The Universal Declaration of Rights of Mother Earth, adopted at the World People's Conference on Climate Change and the Rights of Mother Earth, also recognises "Mother Earth" as a living being (art 1), with certain rights of its own (art 2), and duties that human owe to it (art 3).

[44] *Miglani* (n 6) [62.3].

[45] 'Read: G.D. Agarwal's Third and Final Letter to PM Modi on Saving the Ganga' *The Wire* (11 October 2018) <https://thewire.in/rights/read-gd-agarwal-final-letter-narendra-modi-saving-ganga> accessed December 9, 2018.

[46] Guardians and Wards Act (n 43) s 20.

[47] ibid, s 39(g).

These potential conflicts make such officials unsuitable for a dual role as Ganga's guardians.

Although Stone envisaged a model of guardianship to confer rights on natural entities,[48] his was distinct from the kind formulated by the High Court. In his model, a "friend" of the relevant natural entity had to make an application to be appointed as its guardian.[49] Stone was wary of state functionaries being appointed as guardians of natural entities, given their existing functions under law and their vulnerability to political diktat.[50] In New Zealand, however, the first two appointees as guardians of the Whanganui River were not presently serving as government officials and had sufficient experience in dealing with complex environmental and cultural issues.[51] The High Court may well have chosen persons or organisations with sufficient experience in handling environmental issues as guardians of Ganga, and its decision to appoint incumbent government actors as the guardians of Ganga without any consideration for their knowledge of or experience in handling environmental issues may not be the best model of guardianship to rejuvenate the river.

There is another significant defect from which the judgments suffer: the guardians have not been granted any real powers to review the decisions of the government of the State of Uttarakhand pertaining to Ganga. The authority of Ganga's guardians is further compromised by their lack of financial autonomy to act in Ganga's best interest. To address the damage caused to a natural entity, Stone proposed setting up a trust fund to be administered by the guardian of the entity.[52] As mentioned earlier, the guardian body in New Zealand has a dedicated fund at its disposal to act in best interest of the Whanganui River. In *Miglani*, the High Court records the disbursal of ₹2 billion by the central government for the rejuvenation of Ganga.[53] That expenditure, however, is ex gratia and the government has not yet committed a definite sum of money for the rejuvenation of Ganga, making the guardian body financially dependent on the central government to undertake activities in the best interest of the river.

The High Court could have directed the Uttarakhand government to commit a certain sum of money to the cause of protection of Ganga in the form

[48] Christopher Stone, 'Should Trees Have Standing? – Towards Legal Rights for Natural Objects' (1972) 45 Southern California Law Review 450, 464.

[49] ibid.

[50] ibid, 473.

[51] New Zealand Government, 'First Te Pou Tupua appointed [4/9/17]' *Scoop* (12 September 2017) <http://www.scoop.co.nz/stories/PA1709/S00132/first-te-pou-tupua-appointed-4917.htm> accessed 8 March 2019.

[52] Stone (n 58) 480.

[53] *Miglani* (n 6) [48].

of a trust fund. Indeed, such directions have been given by the High Court in other environmental matters. The Madras High Court recently directed the Tamil Nadu State Legal Services Authority to establish an "Environmental Fund" for the planting of trees and the maintenance of water bodies in the State of Tamil Nadu.[54] The Uttarakhand High Court therefore had jurisdiction to order the creation of a comparable fund.

The High Court in *Miglani* recognised the need to allow people living on the banks of Ganga to have their voices heard.[55] To that end, it allowed the Chief Secretary of the State of Uttarakhand to co-opt up to seven persons to represent local communities.[56] Certainly, allowing for such participation in matters concerning human livelihoods is a key facet of procedural justice,[57] and the High Court's effort must be lauded on that count. The community's right to representation, however, is exposed as inadequate when probing questions are posed. First, the High Court conferred on the Chief Secretary an unfettered discretion to pick these seven representatives at his whim without any principles controlling such selection. Under the principles of administrative law in India, such uncontrolled discretion cannot be exercised by any State official and any action taken in exercise of such discretion is liable to be struck down.[58] Secondly, the nature of community representation — that is, whether community views are binding on the guardian body — remains unclear. If community voice was binding then the High Court or the Supreme Court may be moved for issuance of a writ of mandamus for the guardian body to implement the recommendation of the seven local representatives.[59]

E. Ganga's legal personality and its conflict with the secular values of the Indian Constitution

The preamble of the Constitution declares India to be a "sovereign, socialist, *secular*, democratic republic".[60] Secularism has been held to be a part of the basic structure of the Indian Constitution,[61] meaning that the commitment to this value

54 'In a first, Madras High Court sets up "Environmental Fund"' *DownToEarth* (29 June 2018) <https://www.downtoearth.org.in/news/environment/in-a-first-madras-high-court-sets-up-environmental-fund--60976> accessed 20 March 20, 2019.

55 *Miglani* (n 6) [53].

56 ibid [62.4].

57 Andre Silveira, 'The multiple meanings of justice in the context of the transition to a low carbon economy' (2016) University of Cambridge Institute for Sustainability Leadership 2/2016, 6 <https://www.cisl.cam.ac.uk/resources/publication-pdfs/the-multiple-meanings-of-justice-in-the-context-of.pdf> accessed 8 March 2019.

58 *In re Delhi Laws Act, 1912* AIR 1951 SC 332 (Supreme Court of India).

59 Constitution of India (n 13) arts 32 and 226.

60 ibid, preamble (emphasis added).

61 *Bharati v State of Kerala* (1973) 4 SCC 225 (Supreme Court of India).

cannot be altered or abrogated even by Parliament through law. The judicial system is bound to uphold secular values and ensure that its actions are not tainted with religious overtones. A major factor that has weighed disproportionately in the High Court's reasons for granting legal personality to Ganga is the "religious importance"[62] and the "piety"[63] of Ganga. While the river is highly venerated among the followers of the Hindu religion, the focus of the Court on its religious significance may be vulnerable to misuse by parochial, communal interests.[64]

This is a cause for special concern given the rise of Hindu nationalism has placed the Bharatiya Janata Party (BJP) in central government and that of the State of Uttar Pradesh, which enjoys the longest course of the Ganga flowing through its territory. The High Court should have been cognisant that excessive reliance on Ganga's religious importance had the potential to marginalise its non-religious ecological importance. Indeed, there have been instances in the history of Indian struggles in which environmental rhetoric has been co-opted by religious or nationalist organisations to serve their own agendas and interests.[65] A germane example is when the Vishwa Hindu Parishad, a Hindu nationalist organisation, made appeals for the demolition of Tehri Dam on Ganga by drawing justificatory analogies with the 1992 demolition of the Babri Mosque in Ayodhya, which had triggered riots and deaths across India.[66] Muslims were openly shunned from the organisation's protests against the Tehri Dam, as the Vishwa Hindu Parishad claimed that Muslims were the "unworthy, bad sons" of Ganga.[67] Given the susceptibility of Ganga's iconography for misuse, the High Court would have been

62 *Salim* (n 6) [17].

63 *Miglani* (n 6) [4].

64 Erin L O'Donnell, 'At the Intersection of the Sacred and the Legal: Rights for Nature in Uttarakhand, India' (2017) 30(1) Journal of Environmental Law 135, 140–141.

65 See generally Mukul Sharma, *Green and Saffron: Hindu Nationalism and Indian Environmental Politics* (Permanent Black, 2011).

66 Mukul Sharma, 'Nature and nationalism' *Frontline* (3 February 2001) <https://frontline.thehindu.com/static/html/fl1803/18030940.htm> accessed 20 March 2019. A similar Hindu-nationalist brand of environmentalism can be seen with respect to the Vrindavan Forests located in Mathura district. The Vrindavan Forest Revival Project, later rechristened as the Vrindavan Conservation Project, identified the main causes of the degradation of the Vrindavan forests as the abandonment of Hindu values and the centuries of "Muslim and British Rule": Mukul Sharma, 'Saffronising green' (2002) <http://www.india-seminar.com/2002/516/516%20mukul%20sharma.htm> accessed 20 March 2019.

67 Sharma, 'Nature and nationalism' (n 76).

wise not to rely disproportionately on Ganga's religious piety as a reason to confer legal personality.

F. Conclusion

When we stack these factors, it becomes apparent that the judgments of the High Court do not add sufficient substance to Ganga's legal personality, despite paying lip service to the intrinsic value of its life. The personality crafted by the High Court relies heavily on the grace of government officials to act in Ganga's best interest, which might not always be the case due to the likelihood of conflicted interests. Indeed, Ganga's guardians lack financial and executive autonomy from the governmental apparatus, severely curtailing their ability to take positions the government does not approve of. The rich international jurisprudence on the protection of rights of nature and legal personhood has also not been engaged with in any seriousness despite the High Court's awareness of the same, further doing a disservice to Ganga's protection. The religious overtones present in the High Court's reasoning to confer personality on Ganga can be easily co-opted by fanatical forces to turn an environmental struggle into a parochial, communal struggle on sectarian grounds. The High Court's declaration of Ganga's legal personality rings hollow in face of these limitations in its reasoning. Nonetheless, a combined reading of these judgments with those of the Supreme Court of India, which recognise ecocentrism as a principle of law, allows for a substantive elaboration of Ganga's legal status.

IV. Ecocentrism and the Future of Ganga

While the judgments of the Uttarakhand High Court are flawed from an environmental perspective for the reasons given above, they nonetheless indicate a discursive shift in the emerging jurisprudence on environmental law in India, which has previously focused on human interests in the environment of a decidedly anthropocentric nature.[68] The Indian courts are beginning to acknowledge the intrinsic value of nature, however, and the idea of ecocentrism as a principle of law is emerging in the Indian legal system. Ecocentrism is a principle of environmental ethics that recognises the intrinsic value of life of all beings and ecosystems that inhabit and function on earth.[69] An ecocentric approach is starkly opposed to the anthropocentric approach, which recognises only the intrinsic value in human

[68] Zafar Mahfooz Nomani, 'The Human Right to Environment in India: Legal Precepts and Judicial Doctrines in Critical Perspectives' (2000) 5 Asia Pacific Journal of Environmental Law 113, 128.

[69] Andrew Brennan and Yeuk-Sze Lo, 'Environmental Ethics' *The Stanford Encyclopedia of Philosophy* (Winter Edition, 2016) <https://plato.stanford.edu/entries/ethics-environmental/> accessed 20 March 2019.

beings and sees nature and natural entities merely as means to serve human ends.[70] Indeed, the High Court in *Miglani* recognised the intrinsic value of Ganga and its ecosystem as a major reason for why legal personality should be conferred.[71]

This mere mention of Ganga's intrinsic value, and its recognition as a living entity,[72] may be rendered conceptually coherent and legally enforceable if the High Court judgments are read alongside recent Supreme Court judgments in the Gir lions case and the Jallikattu case.[73] This section argues that the question of Ganga's rights, especially its right to life, may be firmly rooted in juridical discourse that allows for meaningful protection by creatively reading the High Court judgments in light of the ecocentric jurisprudence that has been laid down by the Supreme Court.

Ecocentrism and Ganga's Right to Life

The judgment in the Gir lions case began with the Supreme Court declaring that the issue of translocation of Asiatic lions from a wildlife sanctuary in Gujarat to another in Madhya Pradesh must be decided in line with ecocentric principles.[74] The Supreme Court applied the "species best interest standard" to come to the conclusion that a pride of lions must be translocated to protect the population of the lions from epidemic breakouts and other threats to their survival.[75] That standard was also applied to ban the bull-taming sport of Jallikattu, as it was considered to be against the well-being of the bulls that were made to participate and because the sport caused infliction of unnecessary pain and suffering.[76] The crux of the judgment in the Jallikattu case is its acknowledgement of the intrinsic value of the lives of bulls and its critique of the anthropocentric bias in law and in human practices that refuse to acknowledge animals' right to be treated with dignity and respect.

The Jallikattu case has fascinating legal implications bearing directly on the rights of Ganga. One paragraph from the judgment implies that the right to life extends to living non-human entities in light of its expansive interpretation of "life" under Article 21 of the Indian Constitution.[77] A combined reading of

[70] Ibid.
[71] *Miglani* (n 6) [50].
[72] ibid [62.2].
[73] *Centre for Environment Law v Union of India* WP (Civil) No. 337/1995 (Supreme Court of India) (Gir lions case); and *Animal Welfare Board of India v Nagaraja* Civil Appeal No. 5387/2014 (Supreme Court of India).
[74] Gir lions case (n 83) [1].
[75] ibid [40].
[76] Jallikattu case (n 83) [46].
[77] ibid [62].

the Jallikattu case with the *Miglani* judgment, which recognised Ganga as a living entity, implies that Ganga has a constitutional right to life. Notwithstanding the legal logic of this result, it remains a daunting task to determine what precisely a right to life might mean for Ganga in practice. Moreover, understanding and enforcing Ganga's right to life presupposes an understanding of what it means for Ganga to have life, a deeply complex question unto itself. To effectively enforce Ganga's right to life, it is imperative that a standard be adopted against which Ganga's life may be measured.

Perhaps a workable albeit imprecise metric of Ganga's life could be its ecological flow. Ecological flow connotes the minimum quantity and quality of water that must flow through a river at any given point of time for the continued sustenance of the river and its ecosystem and of the human life that it supports.[78] The central government has recently notified the ecological flow that ought to be maintained in different stretches of Ganga in different seasons, with Namami Gange having the power to revise the quantum of these ecological flows.[79] There is a duty on the designated authorities to ensure that the ecological flow does not fall below the prescribed limits to maintain the "ecological integrity" of the river.[80] Violation of this threshold might be a suitable proxy for the right to life of Ganga; it has been argued that this recent notification may be used to protect the health, and thereby the life, of the Gangetic river system.[81]

Another way Ganga's right to life might be protected is by protecting the right to life of the fauna that lives in, or is dependent on, the river's waters. The Supreme Court in the Jallikattu case extended the protection of the right to life for non-human animals.[82] Given that the Ganges River dolphin, for instance, has been classified as an endangered species by the International Union for Conservation of Nature, it may be argued that its right to a clean environment[83] and to live with

[78] Ben Gillespie, 'What are environmental flows?' (*The River Management Blog*, 28 April 2014) <https://therivermanagementblog.wordpress.com/2014/04/28/what-are-environmental-flows/> accessed 20 March 2019.

[79] Jacob Koshy, 'Centre sets 'minimum river flows' for the Ganga' *The Hindu* (Chennai, 10 October 2018) <https://www.thehindu.com/sci-tech/energy-and-environment/centre-sets-minimum-river-flows-for-the-ganga/article25184144.ece> accessed 20 March 2019.

[80] River Ganga (Rejuvenation, Protection and Management) Authorities Order 2016, para 5.

[81] Ritwick Dutta, 'A notification to ensure ecological flow of rivers' *India Water Portal* (12 December 2016) <https://www.indiawaterportal.org/articles/notification-ensure-ecological-flow-rivers> accessed 20 March 2019.

[82] Jallikattu case (n 83) [62].

[83] *Mehta v Union of India* WP (Civil) No. 13029/1985 (Supreme Court of India).

dignity[84] under Article 21 of the Indian Constitution is violated by the polluting industries, dams and other agents that cause damage to the river.

On a broader level, the Supreme Court in the Jallikattu case based the duty to protect bulls and all other living entities on Article 51A(h) of the Constitution — "the duty of every citizen of India to develop the scientific temper, humanism, and the spirit of inquiry and reform" — by giving the term "humanism" an expansive definition encompassing "inclusive sensibility for our species".[85] Given that Ganga has been recognised as a living entity by the High Court, the inclusive sensibility posited by the Supreme Court might be broad enough to cover not just living organisms like animals but also living systems like Ganga and its ecosystem.

It thus becomes clear that, while the High Court judgments make only a hollow declaration of Ganga's legal personality, richer meaning may be added to those judgments within the broader context of the ecocentric jurisprudence that has been laid down by the Supreme Court. A direct implication of reading the High Court judgments in light of the Supreme Court judgments is that Ganga's right to life under Article 21 of the Indian Constitution can be effectively protected and enforced through the courts. With the adoption of a reasonable metric to gauge Ganga's life — namely ecological flow — and punishing any violations of the notified threshold, or instituting suits in the name of members of the animal kingdom which inhabit Ganga whenever any step is taken that leads to pollution, Ganga's right to life may be guarded against actors contributing to its pollution. On a more metaphysical level, it may be inferred from the above discussion that the duty of Indian citizens to instil humanism within themselves also requires that they recognise the right to life and personality of living systems such as the Ganga, although the justiciable status of that duty falls beyond the bounds of this paper.

V. Conclusion

Given the ongoing degradation of Ganga, despite the death of Professor Agarwal while fasting for its protection, it is imperative that a robust protective mechanism is developed expeditiously. Legal personality is one such mechanism, but there are various problems with the reasoning of the *Salim* and *Miglani* judgments of the Uttarakhand High Court. The judgments do not lay down any framework of enforceable rights for the river Ganga, which renders the declaration of its personhood deficient in substance. The High Court invokes the doctrine of parens patriae to confer personhood on Ganga despite its lack of competence to do so. The High Court thereby casts Ganga as a perpetual minor and denies it full-fledged personhood. It does not engage with other international developments

[84] *Gandhi v Union of India* AIR 1978 SC 597 (Supreme Court of India).

[85] Jallikattu case (n 83) [58].

in the rights-of-nature movement, despite being aware of the same. Ganga's guardians lack the financial autonomy and the powers to review governmental action to act in Ganga's best interest. The disproportionate importance given by the High Court to Ganga's religious piety, moreover, lends itself to the agenda of Hindu-nationalist organisations that may pervert it for their own purposes.

In sum, the judgments of the High Court each make an empty declaration of Ganga's legal personality without fleshing out the particulars for effective enforcement. These declarations may be augmented, however, if the judgments are read creatively in light of judgments of the Supreme Court that have recognised ecocentrism as a principle of law. Such a reading would allow us to identify and protect Ganga's right to life through the courts. With the Ganga Bill now being circulated in select circles, it is imperative for the basis of Ganga's legal personality to be anchored in principles of ecocentrism. This idea of an ecocentric personality for Ganga must be taken to its logical conclusion in the Ganga Bill through statutory recognition of Ganga's right to life and continued existence, tracing it from the provisions of the Indian Constitution, and institutionalise a mechanism targeted at combating the systemic reasons for and source of Ganga's pollution.

While certainly a legislation response to the protection of Ganga would be of much help, the non-availability of the text of the Bill for public discussion is lamented. There is a pressing need to debate and discuss the Ganga Bill and allow for adequate public participation in its formulation, so that a democratic and environmentally and socially just framework may be developed to protect Ganga and its rights. The scant detail known of the Ganga Bill seems only to consolidate under one Act what already exists under other laws; there is little reason to believe that it will lead to a radically different outcome for the alarmingly deteriorating health of the river. It is only when Ganga is properly managed, however, and is endowed with the ability to enforce its own rights through independent agents that it may be said that meaningful legal personhood has been conferred on Ganga. Principally, it must be ensured that the governance mechanism designed to protect Ganga is immune from political interference and rooted in a scientific, ecocentric ethos with protection of Ganga as its primary goal. To provide for anything more than this without a rigorous study of Ganga's ecosystems would be to ignore its varied relationships with communities and geographies, and a disservice to the cause of its preservation and rejuvenation. But if the Ganga Bill considers Ganga's interests rather than those of humans in Ganga and puts in place a robust machinery to defend the ecosystem's interests, then it may be able to do what

programmes and policies of successive governments have repeatedly failed at: turning the clock back on Ganga's rapid degradation.

Christian Divorce Law in Pakistan: Past, Present and Future

Muhammad Zubair Abbasi[*]

Abstract

The law governing the divorce of Christians in Pakistan was enacted through the Divorce Act 1869 during the British colonial period in India. To ensure incremental changes in this law, section 7 of the Act provided that the local courts will apply English divorce law for Christians. In 1981, however, this section was repealed under the regime of Zia ul-Haq through an Ordinance. In 2017, a member of the Christian community challenged this repeal on the ground that it violated constitutionally guaranteed fundamental rights of minorities in Pakistan. The Lahore High Court accepted the petition and declared the repeal of section 7 as unconstitutional. The revival of section 7 has a significant impact on Christian divorce law in Pakistan by providing, *inter alia*, irretrievable breakdown of marriage as a ground for divorce. However, it has led to the anomaly of English divorce law applying to Christian citizens in Pakistan. To remove this anomaly, the Pakistani parliament will have to reform Christian divorce law.

I. Introduction

In Pakistan, the Christian community comprises 1.59 percent (3.29 million) of the total population of 207 million. Despite being the second largest religious community in Pakistan, Christian divorce law in Pakistan is based on the statutes which were passed during the second half of the nineteenth century. As a result, Christian couples were forced to live in dead marriages because irretrievably

[*] Assistant Professor, Lahore University of Management Sciences (LUMS); zubair.abbasi@lums.edu.pk.

broken-down Christian marriages could not be dissolved under the law. All this, however, changed in 2017 when the Chief Justice Mansoor Ali Shah of the Lahore High Court delivered a landmark judgment recognizing irretrievable breakdown as a valid ground for divorce under Christian divorce law in Pakistan.

In this article, I analyze the judgment in *Ameen Masih v Federation of Pakistan*[1] to assess its implications for Christian divorce law in Pakistan. This article is divided into three parts. Part 1 traces the historical development of Christian divorce law during colonial and post-colonial periods and illustrates how General Zia ul-Haq's amendment of the Divorce Act 1869, coupled with the promulgation of Islamic criminal law against extra-marital sex (*zina*), condemned Christian couples to live in dead, and often abusive, marriages. Part 2 examines the judgment of the Lahore High Court, in which the Court recognized irretrievable breakdown as a valid ground for divorce. Part 3 explores the implications of this judgment for the Christian community in Pakistan. This article points out that although the judgment has provided the irretrievable breakdown of marriage as a valid ground under Christian divorce law, it has left the Christian citizens of Pakistan to be governed under the provisions of foreign English divorce law. The Pakistani parliament will have to remove this anomaly by expanding the scope of divorce under Christian law through a statute.

II. Christian Divorce Law In Pakistan: Historical Context

In Christian law, marriage is a sacred covenant which must not be broken. The New Testament provides very narrow grounds for divorce. The Gospel of Matthew states: 'whoever divorces his wife, except for immorality, and marries another woman commits adultery'.[2] The divorce laws of many Christian countries, therefore, provide only limited grounds for divorce. In the United Kingdom, the Church of England, as a separate entity from the Catholic Church, emerged in part due to the Pope's refusal to grant an annulment to King Henry VIII from his marriage to Catherine of Aragon because she had been unable to give birth to a son and heir to the King.[3] The Divorce Reform Act 1969 provided the ground of irretrievable breakdown of the marriage for divorce under English law. Further development and consolidation of English divorce law was undertaken through the Matrimonial Causes Act 1973 which became the primary law governing marriage

[1] PLD 2017 Lahore 610.

[2] Matthew 19:7-9, *The Holy Bible*, containing the Old and New Testaments, King James Version (New York: American Bible Society 1999) 2062.

[3] David G Newcombe, *Henry VIII and The English Reformation* (Routledge 1995) 12.

and divorce law in England and Wales. Despite these developments, a unilateral right to no-fault divorce is not available under English law.[4]

In colonial India, the law governing Christian divorce was first enacted in the form of the Divorce Act 1869. It provided limited and unequal grounds for dissolution of marriage for both the husband and the wife. Under section 10 of the Act, a husband could petition the court to grant a divorce by accusing his wife of adultery. A woman could also petition for divorce under the same ground of adultery, but additionally, she had to prove that the husband had changed his religion from Christianity or committed bigamy, or rape or sodomy or bestiality, or cruelty or desertion, without reasonable excuse, for two years or more.[5] During the British colonial period in India, this statute was amended to update the law with the passage of time to accommodate societal changes. The most significant change occurred in 1912 when section 7 was added to the Act. Under this section, a relief of dissolution of marriage could be given in accordance with the contemporary principles of divorce law as applied in England. Section 7 provided:

> Court to act on principles of English Divorce Court: Subject to the provisions contained in this Act, the Courts shall, in all suits and proceedings hereunder, act and give relief on principles and rules which, in the opinion of the Courts, are as nearly as may be conformable to the principles and rules on which the Court for Divorce and Matrimonial Causes in England for the time being acts and gives relief.

In 1981, however, during the regime of General Zia ul-Haq, section 7 of the Act was repealed through the Federal Laws (Revision & Declaration) Ordinance 1981. This removal of section 7 changed the original spirit of the Act, making it restrictive and ossified its provisions. Since the UK Matrimonial Causes Act 1973 had given the irretrievable breakdown of a marriage as a ground for

[4] In a recent judgment, the UK Supreme Court denied a wife's petition to divorce her husband on the basis of irretrievable breakdown of marriage. *Owens v Owens* [2018] UKSC 41.

[5] Section 10 states: "Any husband may present a petition to the Court of Civil Judge praying that his marriage may be dissolved on the ground that his wife has, since the solemnization thereof, been guilty of adultery. Any wife may present a petition to the District Court or to the High Court Division, praying that her marriage may be dissolved on the ground that, since the solemnization thereof, her husband has exchanged his profession of Christianity for the profession of some other religion, and gone through a form of marriage with another woman; or has been guilty of incestuous adultery, or of bigamy with adultery, or of marriage with another woman with adultery, or of rape, sodomy or bestiality, or of adultery coupled with such cruelty as without adultery would have entitled her to a divorce mensa et toro, or of adultery coupled with desertion, without reasonable excuse, for two years or upwards. Every such petition shall state, as distinctly as the nature of the case permits, the facts on which the claim to have such marriage dissolved is founded."

divorce, the repeal of section 7 foreclosed a ground for divorce which should have been available to Christians in Pakistan just as it was available in English Divorce and Matrimonial Causes Court. With many Christians unwilling to accuse their spouses of adultery, or unable to prove such a charge to get divorce decrees, the law trapped them in their marriages.

The impact of the repeal of section 7 of the Act was exacerbated because of the promulgation of the Offense of Zina (Enforcement of Hudood) Ordinance 1979 which made extra-marital sex a criminal offense. Simultaneously, the Offense of Qazf (Enforcement of Hudood) Ordinance 1979 declared false accusation of *zina* (extramarital sex) as a penal offense. These legal developments had far-reaching consequences for a divorcing Christian couple whose personal dispute about the dissolution of marriage might have led to criminal sanctions either for one or both of them. In this way, after the repeal of section 7 of the Act, a Christian marriage could not be dissolved without leading to criminal penalties under the Hudood Ordinances. This led many Christian women to convert from Christianity in order to escape from their bad marriages. In several reported cases,[6] Christian wives claimed dissolution of their marriages upon their conversion to Islam.

The requirement to prove adultery for getting a decree of dissolution of marriage poses its own challenges. In *Parveen Amanual v Additional District Judge-III*,[7] it was held that a claim for dissolution of marriage should be based on concrete facts supported by reliable and cogent evidence and not mere allegations. The court cannot dissolve a marriage or grant separation on mere assertion that the wife is not willing to live with her husband. In *Mushtaq v Fareeda*,[8] after evaluating the allegations of adultery by a Christian husband, the court refused to confirm the decree of the dissolution of marriage. This meant that a wife, who was accused by her husband of adultery, was still required to live with him. The irretrievable breakdown of their marriage was not recognized as a valid ground for dissolution of marriage. The spouse must prove the allegations of adultery through corroborative evidence either through witnesses or by relying on surrounding circumstances for the decree of dissolution of marriage. Historically, the courts have held a high standard of proof for matrimonial offenses. In *Inayat Bibi v Harbans Lal*,[9] the court observed that the standard of proof for matrimonial offenses is as stringent as in criminal cases i.e., the case must be proved beyond reasonable doubt. Similarly, in *Anges Jacintha*

6 *Salamat Ali v The State* 1989 PCrLJ 978; *Tariq Masih v The State* 2004 PCrLJ 1017; *Afzal Masih v The State* 2004 MLD 970; *Aftab Ahmad v Judge Family Court* 2009 MLD 962.

7 PLD 2009 Lahore 213.

8 CLC 1979 Karachi 457.

9 PLD 1966 Pesh 13.

Irene v Augustus Simon D'Souza, [10] the court held that an inference of adultery cannot be drawn lightly from evidence of association coupled with opportunity. In this case, the court refused to accept as evidence of adultery the fact that the petitioner's husband returned very late from the house of his alleged mistress.

Besides the fact that many petitioners seeking divorce must prove matrimonial offenses according to a standard that is prohibitively difficult, it is worth noticing that adultery is a criminal offense in Pakistan. General Zia ul-Haq introduced the process of Islamization of laws in Pakistan. In 1979, he promulgated four Ordinances to enforce Islamic criminal law. The Ordinances covered the offenses of extra-marital sex (*zina*), the false accusation of *zina* (*qazf*), the drinking of alcohol, and theft. The Offense of Zina (Enforcement of Hudood) Ordinance 1979 applies to both Muslims and non-Muslims. Section 4 of the Ordinance defines extra-marital sex (*zina*) as: "A man and a woman are said to commit '*zina*' if they willfully have sexual intercourse without being married to each other." This section does not make any distinction on the basis of religion. By alleging adultery in order to seek a divorce, the petitioner opened up the possibility of the respondent being charged with a criminal offense which was liable to a maximum of death sentence. Since the burden of proof for both matrimonial and criminal offenses is the same, taking such a ground for divorce could have severe penal consequences. Yet, the failure of a spouse to prove adultery might have made him/her liable to punishment under the Offense of Qazf (Enforcement of Hudood) Ordinance 1979 for false accusation of extra-marital sex (*zina*).[11] The impact of the Hudood Ordinances on Christian divorce law is evident from the fact that not a single case of divorce on the allegation of adultery has been reported since the promulgation of Hudood Ordinances in 1979.

III. Judicial Reform of Christian Divorce Law

In *Ameen Masih v Federation of Pakistan*,[12] a Christian man filed a petition before the Lahore High Court. He contended that since their marriage had irretrievably broken down, he wanted to divorce his wife without accusing her of adultery. He argued that under the repealed section 7 of the Divorce Act 1869, grounds for divorce under the UK Matrimonial Causes Act 1973, were available to him in the courts of Pakistan. These grounds included the ground of irretrievable

[10] PLD 1975 Kar 747. In *Francisco Xavier Pinto v Julie Pinto* PLD 1979 Karachi 716, however, the court held that an inference of the commission of adultery may be drawn from the evidence of association coupled with the opportunity to commit adultery and the evidence of an illicit affection.

[11] *Shaukat Masih v The State* 1987 SCMR 1308; *Yousuf Masih and another v The State* 1994 SCMR 2102.

[12] *Ameen Masih* (n 1).

breakdown of marriage. However, this section was omitted through the Federal Laws (Revision and Declaration) Ordinance 1981. Therefore, he petitioned the court to declare the repeal of section 7 of the Divorce Act 1869 as unconstitutional on the basis of Article 36 of the Constitution of Islamic Republic of Pakistan 1973 which provides that the State is bound to protect the legitimate rights and interests of the Christians as the citizens of Pakistan.[13]

The Lahore High Court, in its judgment, traced the history of Christian divorce law. After observing that, historically, divorce has been frowned upon by the church authorities, the Chief Justice Mansoor Ali Shah held:

> In Christian majority countries, although it is public policy to discourage divorce, and not to favour or encourage it, public policy does not discourage divorce where the relations between the husband and wife are such that the legitimate objects of the matrimony have been utterly destroyed. The State is not interested in perpetuating a marriage after all possibilities of accomplishing a desirable purpose of such relationship is gone, or out of which no good can come and from which harm may result. Accordingly, it is the public policy to terminate dead marriages.[14]

Justice Shah pointed out that it is in such situations that a 'no fault' divorce mechanism is provided in many jurisdictions. Such mechanisms or statutes allow for a no-fault divorce, or divorce by consent, in which the parties are not required to prove fault or grounds for divorce other than showing irreconcilable differences or an irretrievable breakdown of the marriage. He observed that divorce laws have been amended to provide for the irretrievable breakdown of marriage as a ground for divorce in many Christian populated countries. In India, for instance, the Indian Divorce (Amendment) Act 2001 was enacted which expanded the scope of divorce by providing for the dissolution of marriage by mutual consent.[15] This amendment was similar to the provisions for 'dissolution of marriage by mutual consent' for Hindus under Section 13-B of the Hindu Marriage Act 1956, Parsis under Section

[13] Article 36 of the Constitution of Islamic Republic of Pakistan 1973 provides: "The State shall safeguard the legitimate rights and interests of minorities, including their due representation in the Federal and Provincial services."

[14] *Ameen Masih* (n 1) 625.

[15] Section 10-A of the Act provides: "Subject to the provisions of this Act and the rules made thereunder, a petition for dissolution of marriage may be presented to the District Court by both the parties to a marriage together, whether such marriage was solemnized before or after the commencement of the Indian Divorce (Amendment) Act, 2001, on the ground that they have been living separately for a period of two years or more, that they have not been able to live together and they have mutually agreed that the marriage should be dissolved."

32-B of the Parsi Marriage and Divorce Act 1936 and persons marrying under the Special Marriage Act 1954.

Before delving into the jurisprudential analysis, Justice Shah dealt with the objections of the religious clergy who opposed the grounds of divorce other than adultery. In order to ensure effective participation of the Christian community, notices were issued to lawyers, parliamentarians, and Christian religious scholars, seeking their views on the issue. Most of the replies supported the "irretrievable breakdown" of marriage as a sufficient ground for divorce under Christian law. However, a few Christian religious scholars opposed the recognition of the ground of irretrievable breakdown of marriage for divorce as an intrusion in Biblical law by the secular courts. This, according to them, would be a direct affront to the religious sensibilities of the Christian community in Pakistan. To this, Justice Shah replied that in the case at hand, the court was concerned with the constitutionality of a state law. He referred to the oath of judicial office and Article 2A of the Constitution to establish that the judges are duty-bound to uphold, apply, and preserve the Constitution to the exclusion of their personal or religious affinities. He held that under the Constitution, it is the duty of a judge to protect the legitimate rights and interests of the minorities and check all laws pertaining to them on the touchstone of minority rights as enshrined in the Constitution. Justice Shah held:

> This Court is only to judicially review the existing State law on the yardstick of constitutional values and fundamental rights guaranteed to the minorities-cum-citizens of this country under the Constitution. Nothing else. The apprehension of the clergy that this Court is deciding against the teachings of the Holy Bible, is unfounded, as this court is doing no such thing. This Court is simply examining the constitutionality of the provision of the impugned Ordinance whereby section 7 of the Act has been deleted. If the Christian clergy are unhappy with the law, they can approach the Parliament for its revision. Therefore, this case is not about examining the canonical or Biblical law but about assessing the legality and constitutionality of item 7(2) of the Second Schedule of Federal Laws (Revision and Declaration) Ordinance, 1981.[16]

Justice Shah also referred to the international obligations of Pakistan under international covenants such as the Universal Declaration of Human Rights (UDHR) and the International Covenant on Civil and Political Rights (ICCPR),

[16] *Ameen Masih* (n 1) 624.

which require the protection of minority rights, as well as the domestic obligations of Pakistan under the Constitution which provides for the protection of the rights and interests of minorities. Keeping these obligations in mind, Justice Shah held:

> The wisdom and experience behind the liberalization and emancipation of the Christian Divorce law around the world has been the protection of the right to a happy family life and right to dignity of a human being, who cannot be left chained to a dead marriage forever or forced to convert to another religion just to be released of the bondage of an unhappy marriage.[17]

To allow such a situation would "limit the choice of a person to divorce and force a person to lead an unhappy and an oppressive life unless he or she can prove the charge of adultery against the spouse."[18] According to Justice Shah, compared with the divorce rights of the Christians in the contemporary world, the limited grounds of divorce under the Divorce Act 1869 amounted to discrimination against the Christian minority in Pakistan.[19] Justice Shah observed that the right to family life is based on human dignity which demands that the State must establish a legal system that recognizes the right of every person to create a familial relationship according to one's desires.[20] The impugned amendment limited an individual's freedom to seek divorce, thus perpetuating dead marriages which impair the quality of life and curtail the liberty of individuals by forcing them, against their wills, to live unhappy family lives.[21]

Justice Shah held that the impugned Ordinance not only violates the fundamental rights of the Christians, but is also inconsistent with the Principles of Policy as enshrined in the Constitution. He held that even though these principles are not enforceable as positive fundamental rights, this does not mean that the government can enact laws or make policies which are in direct violation of the guidelines provided therein. He thus held that Article 30 of the Constitution does not protect a law which is inconsistent with the Principles of Policy and therefore

17 ibid 636.
18 ibid.
19 ibid 637.
20 ibid 636.
21 ibid.

any such law can be declared unconstitutional for violating constitutional values.[22] In this context, he observed:

> Principles of Policy provide the constitutional aspirations, goals and mission statement for the State of Pakistan. It is a constitutional obligation of the State and its organs and authorities to synchronize with and promote these Principles. These Principles nourish the roots of our democracy and help actualize and fertilize our constitutional values. They are our roadmap to democracy and ensure that the State remains on course to achieve social, economic and political justice. The State or any authority may take time to achieve the said constitutional aspirations due to the non-availability of resources but they cannot at any stage or at any cost go against the Principles of Policy.[23]

Based on the foregoing discussion, Justice Shah restored section 7 of the Divorce Act 1869 and declared unconstitutional and illegal item 7(2) of the Second Schedule to Federal Laws (Revision & Declaration) Ordinance 1981 because it violated minority rights and infringed constitutional values, fundamental rights, Principles of Policy, and international human rights conventions.[24]

IV. Implications of the Ameen Masih Case

By holding the repeal of section 7 of the Divorce Act 1869 as discriminatory and therefore unconstitutional, Justice Shah provided a much needed respite to many in the Christian community in Pakistan. Christian spouses no longer have to allege adultery in order to end failed marriages. With that said, the reintroduction of section 7 into the Divorce Act 1869 is not without its problems. Firstly, by holding that section 7 continues to be a part of the Divorce Act 1869, the court has explicitly tied the development of Pakistani law with the 'principles and rules of the Court for Divorce and Matrimonial Causes in England'.[25] When the statute was originally enacted, the State of Pakistan did not exist and what is currently

[22] Article 30 reads: "Responsibility with respect to Principles of Policy. (1) The responsibility of deciding whether any action of an organ or authority of the State, or of a person performing functions on behalf of an organ or authority of the State, is in accordance with the Principles of Policy is that of the organ or authority of the State, or of the person, concerned. (2) The validity of an action or of a law shall not be called in question on the ground that it is not in accordance with the Principles of Policy, and no action shall lie against the State or any organ or authority of the State or any person on such ground."

[23] *Ameen Masih* (n 1) 638-639.

[24] ibid.

[25] Divorce Act 1869, s 7.

Pakistan was part of the British Empire. In that context, the reliance on British legal principles was not unusual. At that time, a British Viceroy ruled India and the highest Court of Appeal was the Judicial Committee of the Privy Council in London. Today, Pakistan is a sovereign country with its own laws and legal principles. To tie the development or application of a law to a foreign court is problematic because of its conflict with the sovereignty of the state and its judiciary under public international law.

The application of English divorce law for Christians in Pakistan may result in Pakistani courts having to awkwardly apply foreign law to its citizens. In 1963, when section 7 was part of the Divorce Act 1869, the Chief Justice of Lahore High Court in, Manzur Qadir, had regarded this section as anachronistic in the case of *Marie Antoinette v Oswald Robert Palmer*.[26] In this case, a Christian couple sought dissolution of marriage. The issue at hand was whether the husband could be ordered to pay the expenses of litigation of the wife while her suit for dissolution of marriage was pending. The court observed that there was no provision in the Divorce Act 1869, which authorized the court to order such payment. The court, however, asked the husband to make such payment to his wife because the English Matrimonial Causes Rules 1937 provided for the payment of such expenses.[27] On appeal, the case went to a larger bench of the High Court headed by the Chief Justice, Manzur Qadir, who after expressing reservations regarding section 7 of the Act, held:

> This provision is an anachronism in the statute book of Pakistan after Independence, I have no doubt. But that is a matter for the Legislature to consider. The plain duty of a Court is to give effect to the intent of the lawmaker irrespective of other considerations. It seems to me clear that it was the intention of the lawmaker that Courts in this country should refrain from giving relief in circumstances in which the English Divorce Court does not give relief and should give relief where that Court would give it.[28]

As a result of the judgment of the Lahore High Court in the *Ameen Masih* case, Pakistani judges will have to apply the divorce law which is 'conformable to the principles and rules' which are applied by the Court for Divorce and Matrimonial Causes in England.[29] The English law of divorce, however, has been reformed a number of times during the past one and a half century. These reforms are

[26] PLD 1963 WP Lahore 200.
[27] PLD 1957 WP Lahore 235.
[28] Marie Antoinette (n 26).
[29] Divorce Act 1869, s 7.

procedural as well as substantive in nature. Under the procedural reforms, English divorce law has been transformed into secular law. This change occurred almost a decade before the promulgation of the Divorce Act 1869 in British India. Under the Matrimonial Causes Act 1857, when the Court for Divorce and Matrimonial Causes was established, the divorce jurisdiction of Church courts was abolished.[30] The jurisdictional secularization of English divorce law did not automatically lead to its substantive reform because of resistance from the clergy and the Church of England. The change in substantive divorce law happened after "the law of divorce and Canon law of indissolubility of marriage were themselves divorce"[31] and breakdown of the marriage became a ground for divorce under the Divorce Reform Act 1969. While English divorce law was a part of Christian divorce law during the second half of the eighteenth century as a result of the introduction of the Divorce Act 1869 in British India, following the procedural and substantive changes, English law no longer remains a part of Christian divorce law.

As was the case in England, the reform of Christian divorce law in Pakistan has been vehemently opposed by some members of the Christian community in Pakistan. Unlike England, Pakistan follows a personal law system where each religious community is governed by its own personal law.[32] The then Punjab Human Rights and Minorities Affairs Minister, Khalil Tahir Sandhu, stated that he believed marriage to be a religious covenant and, therefore, he did not believe in divorce. He was quoted as stating, in a meeting organized to discuss tabling a bill to enact changes in the Divorce Act 1869, "[How] can a court touch the law ordained by our faith?"[33] During the hearing of the *Ameen Masih* case, the then Federal Minister for Human Rights, Senator Kamran Michael, himself a Christian, cited relevant verses from the Bible to support his stance that religious laws could not be altered to bring them in line with fundamental human rights as that would be a violation of religious principles.[34] He opposed the reintroduction of section 7 in the Divorce Act 1869. After the judgment in the *Ameen Masih* case, some members of the Christian community challenged it before the larger bench

[30] J H Baker, *An Introduction to English Legal History* (4th edn, Oxford University Press 2007) 496.

[31] ibid 497.

[32] About one third of the world's population lives under a system of state-enforced religious family law as opposed to a unified national legal code which governs matters such as marriage, divorce, maintenance, and inheritance. Most of these countries are post-colonial states where the colonial rulers employed a pluralistic legal system which was later inherited by these nations. Yuksel Sezgin, *Human Rights Under State-Enforced Religious Family Laws in Israel, Egypt and India* (Cambridge University Press 2013) 1–12, 205–214.

[33] Asif Aqeel, 'Why Divorce is Close to Impossible for Christians in Pakistan' (*Herald Magazine*, 25 July 2016) <http://herald.dawn.com/news/1153471> accessed 18 February 2019.

[34] 'LHC Reserves Ruling on Christian Divorce Law after Leaders Oppose Changes' (*The Dawn*, 21 January 2017). <www.dawn.com/news/1309764> accessed 18 February 2019.

of the Lahore High Court by filing an appeal which has been pending adjudication since 2017. This resistance to the reinstatement of section 7 in the Act is justifiable on the ground that English divorce law has been secularized during the nineteenth century. Therefore, it seems inappropriate to subject the Christian community in Pakistan to that law.

Despite objections to the secular nature of contemporary English divorce law, the reintroduction of section 7 into the Divorce Act 1869 is likely to protect the rights of Christian wives by replacing the redundant provisions of the Act with English divorce law. For instance, section 39 of the Divorce Act 1869 provides the court the power to order the settlement of a wife's property if a dissolution of marriage or judicial separation has occurred due to adultery on the part of the wife. This section provides that the court, "may order such settlement as it thinks reasonable to be made of such property or any part thereof, for the benefit of the husband, or of the children of the marriage, or of both." There is, however, no corresponding provision for the settlement of a husband's property in favour of the wife in the Divorce Act 1869. In the Matrimonial Causes Act 1973, section 24 provides the court the power to order the transfer of a property from one party to another or the outright settlement of a property for the benefit of the other party.[35] This section does not discriminate in its application on the basis of gender, unlike the Divorce Act 1869. Similarly, there have been developments in case law with regards to the financial rights of spouses upon divorce in the United Kingdom that may now be applicable in Pakistan. In *White v White*,[36] a landmark judgment of the House of Lords, it was ruled that in deciding the division of assets, a fundamental criterion should be equality between the parties and not necessarily their respective needs or what assets they brought into the marriage. Lord Nicholls held:

> Equality should be departed from, only if, and to the extent that, there is good reason for doing so…. The need to consider and articulate reasons for departing from equality would help the parties and the court to focus on the need to ensure the absence

[35] This section, in part, states: "(1) On granting a decree of divorce, a decree of nullity of marriage or a decree of judicial separation or at any time thereafter (whether, in the case of a decree of divorce or of nullity of marriage, before or after the decree is made absolute), the court may make any one or more of the following orders, that is to say: (a) an order that a party to the marriage shall transfer to the other party, to any child of the family or to such person as may be specified in the order for the benefit of such a child such property as may be so specified, being property to which the first-mentioned party is entitled, either in possession or reversion; (b) an order that a settlement of such property as may be so specified, being property to which a party to the marriage is so entitled, be made to the satisfaction of the court for the benefit of the other party to the marriage and of the children of the family or either or any of them."

[36] [2001] 1 AC 596.

of discrimination.[37]

This approach, similar to the principle of community/matrimonial property, has not been applied in Pakistan and would be a major development in Christian divorce law which has only been marginally reformed in more than one and a half century.

After the judgment in the *Ameen Masih* case, the government proposed the Christian Marriage and Divorce Bill, 2017.[38] The Federal Minister for Human Rights, Dr. Shireen Mazari, has tabled a new Christian Divorce Bill to protect the rights of Christian minorities in Pakistan.[39] The enactment of this law is long overdue and has become necessary in the aftermath of the judgment of the Lahore High Court in the *Ameen Masih* case.

V. Conclusion

In light of the foregoing discussion, it is evident that the Lahore High Court judgment in the *Ameen Masih* case is a landmark development regarding Christian divorce law in Pakistan. As a result of this judgment, irretrievable breakdown of marriage is available as a valid ground for the dissolution of a Christian marriage. The even more revolutionary impact of this judgment has been the removal of gender discriminatory provisions of the law applicable to Christian divorce under the Divorce Act 1869. It is likely that divorced women will now be entitled to matrimonial property because of the application of English divorce law upon Christian divorce proceedings in Pakistan.

The positive implications of the judgment in the *Ameen Masih* case, however, have to be taken cautiously for a number of reasons. Firstly, the Christian community in Pakistan holds divergent views regarding divorce given the historical treatment of a Christian marriage as a sacrament dissolvable only upon death. Therefore, despite the benevolent attitude of the Lahore High Court towards the minority community of Christians in Pakistan, the majority of them may not accept the application of secularized English divorce law. Secondly, the English divorce law itself requires reform because unlike the law in a number of jurisdictions, it does not recognize a unilateral right to no-fault divorce. Thirdly, the application of a foreign divorce law to Christians in Pakistan is anomalous because it not only

[37] ibid.

[38] 'Christian Marriage, Divorce Bill to be Sent to Law Ministry for Vetting' (*The News*, 26 October 2017) <www.thenews.com.pk/print/239775-Christian-marriage-divorce-bill-to-be-sent-to-law-ministry-for-vetting> accessed 11 February 2019.

[39] Myra Imran, 'New Christian Divorce Bill to be Formulated' (*The News*, 14 September 2018) <www.thenews.com.pk/print/368174-new-christian-divorce-bill-to-be-formulated> accessed 11 February 2019.

contradicts the notion of the sovereignty of state laws, but also leaves the Christian minority to be governed by the law which is beyond the control of the legislature and judiciary in Pakistan. Therefore, as a result of the judgment in the *Ameen Masih* case, the parliament in Pakistan will have to pass the Christian Divorce Bill to expand the scope of Christian divorce under Pakistani law.