



2022 - 2023 VOLUME VIII

Three
Verulam
Buildings
Barristers

ONE ESSEX COURT

THE
CONSTITUTION
SOCIETY

6KBW
COLLEGE HILL

serle court



LSE Law



THE LONDON SCHOOL
OF ECONOMICS AND
POLITICAL SCIENCE ■

Houghton
St Press

LSE Law Review



Volume VIII

LSE Law Review

Editorial Board 2022/23

Editor-in-Chief

Bennett Jones

Deputy Editor-in-Chief

Darshana Kumar

Articles Editors

Jason Leung	Raya Riyaz	Leonardo Antonelli
Nicola Alroe	Joaquin Caprarulo	Evgeni Kostevitch
Nhan Pham-Thanh	Swantika Kumar	Grace Wang
	Elyse Barg	
	Sarah Lok	

Notes Editors

Olga Efraimidou	Tevž Sitar	Xinmiao Wang
Jackie Veeger	Sarah Ahmad	Derek Qu
Fee Robinson	Alessandro Copola	Ovi Cheung
	Kofo Boboye	

Publications Editors

Sophia Hassel	Surabhi Bansal
Lukrecija Urnevičute	Andra Sipos
Iyad Mohamed	Demi Adedoyin-Adeyinka
	Vincent Chen

Liaison Officers

Kayan Sayeed
Rishma Mehta
Scarlet Poon

Junior Editors

Alan Zhao	Anna Kordellidou	Damla Karabay
Daniel Beech	Emma Chew	Eunsoo Shin
Faith Tan Gin Ann	Gillian Choy	Jeffrey Tao
Laura Lu	Violette Maris	

Design Editor

Amit Pandya

IT Editor

Will Liu

Summer Board

Cham Mikhaeil	Ilsu Ari	Olivie Holcova
Rahul Wijewardane	Ritchard Ninov	Selena Kishek
Claire Li	Ipsa Yadav	Jack Johnson
Joshua Paul	Evelynn Law	Nicole Luk
Rachel Lye	Oliver Lai	

Special Thanks To The LSE Law School

The views expressed by the contributors are not necessarily those of the LSE Law Review Editorial Board. While every effort has been made to ensure the accuracy and completeness of information contained in this journal, the Editorial Board cannot assume responsibility for any errors, inaccuracies, omissions, or inconsistencies contained herein.

This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>. This license allows for copying and distributing the work in any form and to remix, transform and build upon the material, including for commercial purposes, providing author attribution is clearly stated. Note, copyright restrictions may apply to some images; see source information for individual licensing terms, where they differ.

The LSE Law Review uses the Fourth Edition of Oxford University Standard for Citation of Legal Authorities (OSCOLA).

Published in the United Kingdom by the LSE Law Review LSE Student Union Law Society
1 Sheffield Street, London WC2A 2EY Email: editorialteam@lselawreview.com

In affiliation with the London School of Economics Law School.

Text © LSE Law Review and the original authors. Images © LSE Law Review and the individual authors, or copyright holders attributed in the source information.

Cover Design by Amit Pandya
Cover Photo by Nigel Stead

The views expressed by the contributors are not necessarily those of the LSE Law Review Editorial Board. While every effort has been made to ensure the accuracy and completeness of information contained in this journal, the Editorial Board cannot assume responsibility for any errors, inaccuracies, omissions, or inconsistencies contained herein.

This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>. This license allows for copying and distributing the work in any form and to remix, transform and build upon the material, including for commercial purposes, providing author attribution is clearly stated. Note, copyright restrictions may apply to some images; see source information for individual licensing terms, where they differ.

The LSE Law Review uses the Fourth Edition of Oxford University Standard for Citation of Legal Authorities (OSCOLA).

Published in the United Kingdom by the LSE Law Review LSE Student Union Law Society
1 Sheffield Street, London WC2A 2EY Email: editorialteam@lselawreview.com

In affiliation with the London School of Economics Law School.

Text © LSE Law Review and the original authors. Images © LSE Law Review and the individual authors, or copyright holders attributed in the source information.

Cover Design by Amit Pandya
Cover Photo by Nigel Stead

Proudly Sponsored by:



THE
CONSTITUTION
SOCIETY



ONE ESSEX COURT

6KBW
COLLEGE HILL



Table of Contents

i

Letter From the Editor-in-Chief
Bennett Jones

Summer Issue

1

Houston, We Have a Problem — Jurisdictional Issues of Criminal Law in Outer Space
By Avin Persad-Ford

37

Breaking with Convention: The Supreme Court's Approach to Separability in Kabab-Ji v KFG
By Ryan Henry Ho

52

Neuro-Doping, tDCs and Chess — Are WADA's Regulations under threat?
By Ryan Nathan Au

Winter Issue

120

International Corporate Liability: The Past, Present, and Future of 'Class Action Tourism' in England
By Patrick Kenny

151

Overcoming the dualism: how metaphorically-driven conceptions of the public domain can be a barrier to determining what intellectual property law is and ought to be
By Isabel Hahn

179

Best of Both Worlds?

The interplay between international human rights and the law of armed conflict in cyberspace

By Tatjana Grote

Spring Issue

227

Decentralisation and Recentralisation: An Institutional Analysis of EU Competition Law and the Digital Markets Act

By Ruairí McIntyre

286

A contemporaneous evaluation of the degree of success of the Financial Markets Test Case Scheme in the seminal case of Financial Conduct Authority v Arch Insurance (UK) Ltd and Others (2021) UKSC 1

By Sean Burke

322

Are the Laws Still Poor? Reflections on the Right to Work Limiting the Right to an Adequate Standard of Living

By Chen First

368

Unilateral Extraterritorial Sanctions: The Search for a Jurisdictional Justification under International Law

By Anna Sophie Nagel

396

The Forgotten Sector: Establishing the Need for Stronger Regulation of Livestock Emissions in the European Union – An Emergency Law Perspective

By Katharina Neumann

429

Sovereign Therefore Limited:

The Unconstitutionality of Ouster Clauses for Errors of Law under the British Constitution

By Vincent Lafortune

453

*The Role of Developing Countries in Investor-State Arbitration: Reflections on Tethyan
Copper v Islamic Republic of Pakistan*

(LSE Law Review Prize for Best Blog Submission 2023)

By Sarah Ahmad

Letter from the Editor

On behalf of the Editorial Board, I am proud to present Volume VIII of the LSE Law Review journal.

Building on the work of previous years, the LSE Law Review has undergone a period of phenomenal growth and success this year. The LSE Law Review Blog and Academic Journal received hundreds of submissions from authors at top institutions from across the UK and around the world, and expanded its readership to include over 160 nations. This period of success culminated in our largest-ever Main Journal, which saw the Review publish 13 exceptional pieces across a diverse range of legal areas.

The Role of Developing Countries in Investor-State Arbitration — the prize-winning blog piece this year found within the journal — was one of many that the LSE Law Review Blog team published on its platform. The team thoroughly impressed by cutting their teeth on current and novel issues such as the power of courts to overrule parental decisions in the treatment of their children and the government's recent gender-neutral approach to the statutory definition of domestic abuse.

Building on our strong global presence, the Liaisons team worked incredibly hard to not only reaffirm ties with our existing partners, but also establish new relationships. These efforts have paid substantial dividends in increasing the Review's social media footprint and the number of articles submitted to the Review.

I would first like to thank all of our esteemed sponsors — 3 Verulam Buildings, Serle Court, One Essex Court, 6KBW College Hill, The Constitution Society and the LSE Law School — for their generous financial backing and stalwart support of the Review.

To all of our respected authors, thank you for choosing to publish with us and for your incredible patience and cooperation throughout the entire

editorial process. This publication would, quite literally, not have been possible without your submissions.

To the faculty members of the LSE Law School, your contributions have been invaluable to the success of the Review. Dean Kershaw and Professor Sarah Paterson have both been especially critical in not only actively supporting the organisation as a whole, but in acting as mentors and lodestars to many of the Board's members — myself included.

To the Junior Editors, you have jumped right into the deep end, and your levels of drive and dedication to the Review have exceeded my wildest expectations. You are the future of the Board, and I am confident you will all go on to do great things.

To the Editorial Board members, thank you for your unending diligence and commitment to the success of the Review. Your care towards fostering the development of the Junior Editors has been exemplary, and the academic rigour you have demonstrated this year has been nothing short of outstanding. It has been an absolute pleasure to work alongside such a bright and driven group of individuals in making this year's publication a reality. I hope you have enjoyed your time with the LSE Law Review, and look back fondly in years to come.

To the Heads of Departments (Kayan, Raya, Sophia, Surabhi and Tevž), you have stepped up to the challenge of leading your teams with skill and finesse. You have put out more fires than I can count, and I am incredibly thankful to have had the opportunity to lead the Review alongside you all.

Lastly, to my Deputy Editor-in-Chief, Darshana, I would like to express my deepest gratitude for your unwavering dedication, resilience, and support through thick and thin. I could never have done it without you. You helped actualise our vision for this year's Board and publication, and for that, I can only say one thing: thank you.

Bennett Jones
Editor-in-Chief 2022-23
LSE Law Review Editorial Board

SUMMER ISSUE



2022

Houston, We Have a Problem — Jurisdictional Issues of Criminal Law in Outer Space

Avin Persad-Ford*

ABSTRACT

This paper argues that the current means of prescribing criminal jurisdiction in outer space are inadequate. The Outer Space Treaty and the Intergovernmental Agreement — the main international law instruments that prescribe criminal jurisdiction in outer space — fail to account for many potentially common scenarios in outer space. It would also be impractical to fully transplant the means of prescribing criminal jurisdiction on Earth to apply in outer space, too. This paper argues that the best means of prescribing criminal jurisdiction in outer space is to amend Article VIII of the Outer Space Treaty to allow for a hierarchy of criminal jurisdictions. In descending order of priority, the hierarchy of jurisdictions should be based on territoriality, active nationality, passive personality, universality, and protectivity. Such a hierarchy would respect the underlying principles governing human activity in outer space, while providing for greater certainty as to which countries' criminal laws apply in outer space, and under which scenarios.

*LLB, UCL '21. Articling candidate at Howard C. Cohen & Associates.

INTRODUCTION

In August 2019, media outlets reported that NASA astronaut Anne McClain allegedly accessed Summer Worden's — her ex-wife's — bank account without Ms. Worden's authorisation while Ms. McClain was on the International Space Station ('ISS'). Dubbed 'the world's first space crime',¹ it later transpired that Ms. Worden possibly fabricated her account and is now herself under investigation for lying to the United States' ('US') federal authorities.² Still, the incident raises a question about criminal law in outer space that, it is submitted, has not been satisfactorily answered to date, though not for a lack of trying.³ Namely, if a criminal offence took place in outer space, which nation's criminal jurisdiction should apply?⁴

For Ms. McClain and Ms. Worden, it would have been simple. Given that Ms. McClain and Ms. Worden are both US nationals, the ISS' rules regulating jurisdictional disputes would have granted the US criminal jurisdiction (see Section 2(b) below). More difficult jurisdictional issues would have arisen, however, if the alleged crime had taken place on a spacecraft other than the ISS,

¹ Brandon Spektor, 'The World's First Space Crime May Have Occurred on the International Space Station Last Year' (*LiveScience*, 27 August 2019) <<https://www.livescience.com/anne-mcclain-space-crime.html>> accessed 12 January 2022.

² Chelsea Gohd, 'Astronaut Anne McClain's estranged wife charged with lying about alleged 'space crime' (*Space.com*, 7 April, 2020) <<https://www.space.com/astronaut-anne-mcclain-wife-charged-lying-space-crime.html>> accessed 12 January 2022.

³ Recent proposals include: Danielle Ireland-Piper and Steven Freeland, 'Star Laws: Criminal Jurisdiction in Outer Space' (2020) 44 *Journal of Space Law* 44; Caleb Ohmer, 'When Galaxies Collide: Resolving Criminal Jurisdiction Disputes among Nations in Space' (2019) 43 *Journal of Space Law* 344; P.J. Blount, 'Jurisdiction in Outer Space: Challenges of Private Individuals in Space' (2007) 33 *Journal of Space Law* 299; Lee Seshagiri, 'Spaceship Sheriffs and Cosmonaut Cops: Criminal Law in Outer Space' (2005) 28 *Dalhousie Law Journal* 473; Stacey Ratner, 'Establishing the extraterrestrial: criminal jurisdiction and the International Space Station' (1999) 22 *Boston College International and Comparative Law Review* 323; Karen Robbins, 'The extension of United States criminal jurisdiction to outer space' (1983) 23 *Santa Clara Law Review* 627.

⁴ References to 'criminal jurisdiction' in this paper refers to the jurisdiction to prescribe criminal law — as opposed to the jurisdiction to enforce criminal law — except where noted. See Roger O'Keefe, *International criminal law* (Oxford University Press, 2016) 25.

or even in an area of outer space other than the ISS or another spacecraft. Other international laws forbid nations from claiming sovereignty over any area of outer space (see Section 1(a) below). This would force countries to rely on one of the customary international law's extraterritorial means of asserting criminal jurisdiction in outer space (see Section 2 below).

This paper will argue that the best way of prescribing criminal jurisdiction in outer space is to establish a clear hierarchy among concurrent heads of jurisdiction. In descending order of priority, applicable criminal jurisdiction would be based on territoriality, active nationality, passive personality, universality, and protectivity. Although it is generally accepted that there is no hierarchy among the internationally accepted bases of prescribing criminal jurisdiction on Earth,⁵ this paper argues that this consensus should not apply in outer space. Not only do the arguments militating against a hierarchy not apply in outer space, but the proposed hierarchy would also respect the existing practices in outer space and address the limitations of applying each jurisdictional base individually there.

This paper will argue for a hierarchical approach to determining criminal jurisdiction in outer space in three broad sections. First, it will explain why the two principal means to currently prescribe criminal jurisdiction in outer space — the Outer Space Treaty⁶ and the International Space Station Intergovernmental Agreement ('IGA')⁷ — are inadequate as currently drafted. Second, it will discuss why none of the four traditional means of prescribing criminal jurisdiction on Earth — territoriality, nationality, universality, and protectivity — should be extended to be the sole bases of prescribing criminal jurisdiction in outer space. Third, it will canvass the arguments against a hierarchy of prescriptive criminal jurisdiction bases on Earth, and argue why these should not apply in outer space. It will then propose amendments to the

⁵ *ibid.*

⁶ Treaty on Principles Governing the Activities of States in the Exploration and Use of Outer Space, Including the Moon and Other Celestial Bodies (signed 27 January 1967, entered into force 10 October 1967) 610 U.N.T.S. 205 (Outer Space Treaty).

⁷ Agreement Among the Government of Canada, Governments of Member States of the European Space Agency, the Government of Japan, the Government of the Russian Federation, and the Government of the United States of America Concerning Cooperation on the Civil International Space Station, (signed 29 January 1998, entered into force 27 March 2001), 1998 UST LEXIS 303, TIAS No. 12927 (IGA). References to the 'IGA' throughout this paper are to the IGA's 1998 version, unless noted otherwise.

Outer Space Treaty that include a hierarchy of prescriptive criminal jurisdictions, which would incorporate the advantages of the existing law applicable to outer space while addressing the weaknesses of the existing law.

I. HOW INTERNATIONAL LAW CURRENTLY APPORTIONS CRIMINAL JURISDICTION IN OUTER SPACE

The current law for apportioning criminal jurisdiction in outer space has been described as a ‘bucket ... [containing] many different types of rules and regulations rather than as denoting a conceptually coherent single form of law’.⁸ That bucket has two principal instruments: the Outer Space Treaty and the IGA. This Section will argue that, although both instruments in their current form are the result of political compromises, neither can fully account for potentially common scenarios in outer space in which nations may claim concurrent jurisdiction.

While there are other international instruments that are relevant to outer space — namely the Moon Treaty⁹ and the Bogota Declaration¹⁰ — both have limited relevance in international law. No country that has engaged in self-launched human spaceflight has ratified the Moon Treaty. Only eight countries, meanwhile, have ratified the Bogota Convention.¹¹ The eight equatorial countries party to the Bogota Declaration sought to amend the definition of ‘outer space’ in the Outer Space Treaty to exclude the space above those countries’ territories.¹² Those countries instead sought to classify this space, including the geostationary orbits through which satellites pass, as a natural resource.¹³

The problem with the Bogota Declaration was its political impracticality in terms of enforcement. Most satellites pass over those equatorial

⁸ Francis Lyall and Paul B Larsen, *Space Law: A Treatise* (2nd edn, Routledge 2018) 27.

⁹ The Agreement Governing the Activities of States on the Moon and Other Celestial Bodies (adopted 5 December 1979, entered into force 11 July 1984) 1363 UNTS 3 (‘Moon Treaty’).

¹⁰ Declaration of the First Meeting of Equatorial Countries (signed 3 December 1976) (‘Bogota Convention’).

¹¹ *ibid* art VI.

¹² *ibid* art II.

¹³ *ibid* art I.

countries' geostationary orbits because of the Earth's oblate shape. The countries in which those satellites are registered — primarily, the US — would have refused to request permission from the equatorial states to fly satellites over them.¹⁴

a) Outer Space Treaty

The Outer Space Treaty only partially apportions criminal jurisdiction in outer space. For one, it is clear that the Outer Space Treaty allows for nations to fully assert jurisdiction over individuals aboard a registered spacecraft in outer space. Article VI states that private actors' activities in outer space require a nation's authorisation and supervision.¹⁵ Article VIII further states that 'ownership of objects launched into outer space ... is not affected by their presence in outer space ... or by their return to Earth'¹⁶ and that 'a State Party to the [Outer Space] Treaty on whose registry an object launched into outer space is carried shall retain jurisdiction and control over such object and over any personnel thereof'.¹⁷ The United Nations 1975 Convention on Registration of Objects Launched into Outer Space¹⁸ ('Registration Convention') then requires launching states to register all space objects launched from their jurisdiction before take-off.¹⁹ For acts occurring outside of a spacecraft or where there are competing claims beyond Article VIII's ambit, the Outer Space Treaty extends international law on Earth to outer space. Article I of the Outer Space Treaty provides that the 'exploration and use of outer space, including the Moon and other celestial bodies ... shall be the province of all mankind'.²⁰ Article II then says that outer space is not 'subject to national appropriation by claim of sovereignty, by means of use or occupation, or by any other means'.²¹ The most

¹⁴ Thomas Gangale, 'Who owns the geostationary orbit?' (2006) 31 *Annals of Air and Space Law* 425, 443.

¹⁵ Outer Space Treaty (n 6) art VI.

¹⁶ *ibid* art VIII.

¹⁷ *ibid*.

¹⁸ Convention on Registration of Objects Launched into Outer Space (adopted 12 November 1974, entered into force 15 September 1976) 28 U.S.T. 695, 1023 U.N.T.S. 15 ('Registration Convention') art I.

¹⁹ *ibid* art II (1).

²⁰ Outer Space Treaty (n 6) art I.

²¹ *ibid* art II.

important is Article III, which says that states may conduct activities relating to the exploration or use of outer space ‘in accordance with international law’.²²

Article III’s reference to international law is important not only to prescribe criminal jurisdiction, but also as a barrier against political pressures for other nations to conduct subsequent prosecutions for the same crime(s). Article III would allow, for example, Article XII of the Draft Code of Crimes against the Peace and Security of Mankind²³ and Article XIV of the International Covenant on Civil and Political Rights,²⁴ both of which enshrine the *non bis in idem* (double jeopardy) principle, to limit such prosecutions among signatories to those two instruments. It is therefore essential that nations settle prescriptive criminal jurisdiction in outer space from the onset.

While the Outer Space Treaty and Registration Convention are useful starting points for apportioning criminal jurisdiction in outer space, their limitations are fourfold. First, the Registration Convention fails to account for the growing number of objects launched into outer space for commercial purposes. The Outer Space Treaty and the Registration Convention were signed when registering spacecrafts and other man-made objects in outer space was simple because only state actors launched such crafts. Those state actors primarily did so for military purposes and political gamesmanship during the Cold War, to which the Outer Space Treaty and the Registration Convention needed to respond.²⁵ In the post-Cold War era, however, more private actors are using outer space for commercial purposes, including ‘spaceflight

²² *ibid* art III.

²³ International Law Commission, ‘Yearbook of the International Law Commission 1996’, vol II/2, art 12.

²⁴ United Nations General Assembly, ‘International Covenant on Civil and Political Rights’ (adopted 23 March 1976, entered into force 23 March 1976) art 14, para 7.

²⁵ Tanja Masson-Zwaan and Roberto Cassar, ‘The peaceful uses of outer space’ in Simon Chesterman, David M Malone, and Santiago Villalpando (eds), *The Oxford Handbook of United Nations Treaties* (Oxford University Press 2019) 182.

participation’,²⁶ resource extraction, communications, and satellite imagery.²⁷ Morgan Stanley estimates that the roughly 350 billion USD space industry (as of 2016) will grow to 1 trillion USD by 2040.²⁸ It may therefore be more difficult to track this new and growing activity, which the current law does not contemplate.²⁹ For example, although the Registration Convention may be ‘crystalised’ international law, only 89% of spacecraft and other man-made objects in outer space were registered as of the date of this paper’s submission,³⁰ and more than 60% of which are privately owned.³¹ As this paper will explain below, there are issues with determining jurisdiction solely based on a spacecraft’s registration.

Secondly, and more problematically, the Outer Space Treaty fails to address competing claims or criminal acts occurring in areas of outer space outside of the spacecraft, instead relying on customary international law to resolve the dispute per Article III. However, not only would one traditional overarching means of prescribing criminal jurisdiction fail to cover many potentially common scenarios in outer space (see Section 2 below), customary

²⁶ NASA HQ, ‘Principles regarding processes for criteria and selection, assignment, training, and certification of ISS (expedition and visiting) crewmembers’ (*Spaceref*, 31 January 2002) <<http://www.spaceref.com/news/viewsr.html?pid=4578>> accessed 26 February 2022.

²⁷ Organisation for Economic Co-Operation and Development, *The space economy at a glance* 2007 (OECD Publishing 2007) 48. Military applications of outer space are beyond this paper’s scope except where explicitly noted. But see Cassandra Steer & Matthew Hersch, *War and Peace in Outer Space: Law, Policy, and Ethics* (Oxford University Press 2021) for a comprehensive review of how military applications in space interact with international law.

²⁸ Morgan Stanley, ‘A new space economy on the edge of liftoff’ (*Morgan Stanley*, 17 February 2021) <<https://www.morganstanley.com/Themes/global-space-economy>> accessed 21 February 2022.

²⁹ See Steven Freeland, ‘Newspace, small satellites, and law: finding a balance between innovation, a changing space paradigm, and regulatory control’ in Md Ahmad & Jinyuan Su (eds), *NewSpace Commercialization and the Law* (Centre for Research in Air and Space Law 2017) 107–123.

³⁰ United Nations Office for Outer Space Affairs, ‘United Nations Register of Objects Launched into Outer Space’ (*UNOOSA*, 25 February 2021) <<https://www.unoosa.org/oosa/en/spaceobjectregister/index.html>> accessed 26 February 2022.

³¹ Therese Wood, ‘Visualizing All of Earth’s Satellites: Who Owns Our Orbit?’ (*Visual Capitalist*, 20 October 2020) <<https://www.visualcapitalist.com/visualizing-all-of-earths-satellites/>> accessed 20 October 2022.

international law generally does not recognize a hierarchy of criminal jurisdictions (see Section 3(a) below).

Third, the Outer Space Treaty's emphasis on criminal jurisdiction being designated to the registrant's nation assumes that determination of a rightful registrant is simple. It is not. Although spacecraft are registered in one jurisdiction, modern spacecraft such as the ISS incorporate modules designed in multiple jurisdictions. The Outer Space Treaty's only response to such a situation is to require that nations resolve their dispute 'in accordance with international law'.³² International law, however, struggles to accommodate competing jurisdictional claims based on territoriality. As will be shown below, similar problems have arisen over competing jurisdictional claims in Antarctica.

The problem is more acute in outer space where collaboration is key. For example, under the 'Dragon Programme' between the European Space Agency ('ESA') and China's Ministry of Science and Technology, the ESA's constituent nations and China work together to build each module that they send into outer space. The Dragon Programme is silent about prescribing criminal jurisdiction over satellites.³³ The various nations who collaborate to assemble one spacecraft may therefore disagree about who ought to be able to register that spacecraft and thus assume criminal jurisdiction.³⁴

There is a further difficulty in that the nation in which the spacecraft modules are assembled is not always the nation from which the completed spacecraft is launched into outer space. For example, satellite company OneWeb is jointly owned by UK and Japanese companies. Prior to sanctions imposed against Russia, OneWeb relied on the Baikonur Cosmodrome — a launchpad that Russia leases from Kazakhstan — to launch its satellites into orbit. The Registration Convention would have normally granted Russia, as the nation who held the lease on the platform from which the satellites would launch, with

³² Outer Space Treaty (n 6) art III.

³³ European Space Agency, 'Enter the dragon: ESA and China's joint Earth watch begins' (*ScienceDaily*, 27 April 2004) <<https://www.sciencedaily.com/releases/2004/04/040427054730.htm>> accessed 1 September 2022.

³⁴ Ohmer (n 3) 367-368. See also Nina Tannenwald, 'Law versus power on the high frontier: the case for a rule-based regime in outer space' (2004) 29 *Yale Journal of International Law* 363, 388.

prescriptive jurisdiction. After the Russian invasion of Ukraine, however, the UK and Japanese governments insisted that Russia cede jurisdiction over the satellites to the UK. Russia refused to do so. As of March 2, 2022, the Russian government revoked its permission for OneWeb to use the Baikonur Cosmodrome because of this jurisdictional dispute.³⁵

Fourth, the Registration Convention's attempt to balance political realities undermines its attempt to provide legal certainty to the Outer Space Treaty's Article VIII. Under Article VIII, non-registrant nations may maintain competing jurisdictional claims because the Article only refers to 'jurisdiction', not 'exclusive jurisdiction'.³⁶ Those nations with competing claims are therefore not defeated *prima facie* simply by virtue of Article VIII prescribing criminal jurisdiction to the state of registration.

The Registration Convention would fail to resolve those competing claims. Although the Registration Convention requires launching states to register all space objects launched from their jurisdiction before take-off,³⁷ it simultaneously allows partner states to negotiate separate agreements for jurisdiction and control over individual launches.³⁸ While the Registration Convention is not binding in international law because it is merely a General Assembly Resolution, as Lee Seshagiri argues, 'the ongoing practice of state registration of space objects may have crystallized the convention into customary international law'.³⁹ The OneWeb example above illustrates this. Despite there being four nations with potentially competing jurisdictions, the default position was that Russia — as the nation who controlled the platform from which the satellites would launch — had prescriptive jurisdiction as a result of a contractual agreement between Russia, Kazakhstan, the UK, and Japan.⁴⁰ Likewise, when the UK government agreed with the Indian government

³⁵ Paul Sandle, 'UK company OneWeb suspends Baikonur launches' (*Reuters*, 3 March 2022)

<<https://www.reuters.com/business/aerospace-defense/uk-satellite-company-oneweb-suspends-baikonur-launches-2022-03-03/>> accessed 3 August 2022.

³⁶ Michael Chatzipanagiotis, 'Criminal Issues in International Space Law' (2016) 18 *European Journal of Law Reform* 105, 109.

³⁷ *ibid* art II (1).

³⁸ *ibid* art II (2).

³⁹ Seshagiri (n 3) 483.

⁴⁰ Sandle (n 35).

to launch the satellites from India's Satish Dhawan Space Centre, the default was that India would have prescriptive jurisdiction over those satellites.⁴¹

The Outer Space Treaty's main benefit relates more to its overarching principle of cooperation than its specific means of apportioning criminal jurisdiction. Article IX, for example, requires nations to cooperate with each other if a current or planned space activity harmfully interferes with another nation. Although exercising criminal jurisdiction would not 'cause potentially harmful interference with activities of other States Parties in the peaceful exploration and use of outer space',⁴² it may, as Michael Chatzipanagiotis argues, oblige the state of registry to 'refrain from exercising its jurisdiction' over foreign nationals before consulting with their national state(s).⁴³ As the delegation from India noted during negotiations for the Outer Space Treaty:

When the day comes that men of various nations, through international co-operative efforts, journey into outer space and celestial bodies, many old concepts will have to be forgotten and will, indeed, be out of place in outer space. There should be only one governing concept, that of humanity and the sovereignty of mankind.⁴⁴

If the United Nations General Assembly believed that prescribing criminal jurisdiction in outer space based on the spacecraft's registration would be the best way of exercising the governing concept of 'humanity and the sovereignty of mankind', any means of apportionment should not deviate from such an international consensus except where circumstances require it. As argued, however, such circumstances can exist. Examples include unregistered spacecraft, incidents arising outside of spacecrafts, instances where there are multiple assemblers, or where there are competing jurisdictional claims. In the first two cases, there is no territorial jurisdiction connecting a nation to a crime. In the last two cases, there are potentially competing territorial jurisdictions,

⁴¹ Jonathan Amos, 'OneWeb: UK satellite firm does deal to use Indian rockets' (*BBC News*, 21 April 2022) <<https://www.bbc.com/news/science-environment-61175261>> accessed 2 August 2022.

⁴² Outer Space Treaty (n 6) art IX.

⁴³ Chatzipanagiotis (n 36) 109.

⁴⁴ Bin Cheng, 'The extra-terrestrial application of international law' (1965) 18(1) *Current Legal Problems* 132, 133.

which the Outer Space Treaty cannot resolve in its current form. So, while the present Outer Space Treaty may be a reasonable starting point for apportioning criminal jurisdiction in outer space, it cannot do so comprehensively without amendment.

b) IGA

There are two differences between the Outer Space Treaty and the IGA. First, the Outer Space Treaty covers all of outer space,⁴⁵ whereas the IGA only applies aboard the ISS.⁴⁶ Second, unlike the Outer Space Treaty, the IGA has explicit provisions for allocating criminal jurisdiction,⁴⁷ for which the latter has been termed ‘one of the few positive sources of criminal law in outer space’.⁴⁸ These provisions are still, however, an insufficient basis for adopting a wider means of apportioning criminal jurisdiction in outer space.

The ‘positive source’ to which Professor Blount alludes is deceptively simple. Article V(2) of the IGA says that ‘each Partner shall retain jurisdiction and control over the elements it registers ... and over personnel in or on the [ISS] who are its nationals’.⁴⁹ Article XXII(2), however, then allows the complainant’s nation to exercise exclusive criminal jurisdiction where the alleged perpetrator’s nation ‘fails to provide assurances that it will submit the case to [the perpetrator State’s] competent authorities for the purpose of prosecution’.⁵⁰ For example, if A, a US national, were to assault B, a Russian national, Article V(2) would grant the US default jurisdiction because of A’s nationality. If Russia, however, were sceptical of the US’ ability or desire to refer A for prosecution, then Russia could exercise its criminal jurisdiction through Article XXII(2).

The IGA has three clear advantages. First, it would be simple, on its face, to apply. A Partner State could assert criminal jurisdiction whenever one of its perpetrators was a national, subject to its belief that the nation who registered the area of the ISS in which the alleged incident occurred would not prosecute the alleged offender. Second, Partner and non-Partner nations who use the ISS

⁴⁵ Outer Space Treaty (n 6) art I.

⁴⁶ IGA (n 7) art 1(1).

⁴⁷ Outer Space Treaty (n 6) art VIII.

⁴⁸ Blount (n 3) 312.

⁴⁹ IGA (n 7) art 5(2).

⁵⁰ IGA (n 7) art 22(2).

would be well-accustomed to the IGA's means of apportioning criminal jurisdiction in outer space. This is key because of the large number of individuals who have visited the ISS as a proportion of the total number who have visited outer space. As of November 20, 2020, of the 567 individuals who have reached low Earth orbit, 242 researchers and tourists from 19 nations have visited the ISS.⁵¹ Third, the changes between the 1988⁵² and 1998 IGAs indicate that, although nations acknowledge that their wealthier counterparts may have greater influence over prescribing criminal jurisdiction in outer space, those wealthier countries are willing to compromise. The IGA 1988 allowed the Partner States to exercise jurisdiction over 'the flight elements they provide' and 'personnel in or on any flight element who are their respective nationals'⁵³ — corresponding to subjective territoriality and nationality, respectively (see Sections 2(a) and (b) below). Although notionally a collaboration between Partner States, the IGA 1988 granted the US unprecedented jurisdiction over alleged criminal acts aboard the ISS, likely because the US was the primary provider of equipment and expertise.⁵⁴ When Russia joined the IGA in 1998, the US agreed in exchange to cede its sole criminal jurisdiction not just to Russia, but to the other Partner States, too.⁵⁵

The IGA's simplicity is not without its problems, not only for potentially basing a wider jurisdictional dispute mechanism on the IGA, but because of the IGA itself. First, the IGA is silent on how jurisdiction ought to apply where a crew member's conduct places the rest of the crew in imminent danger and there are several nations with competing claims over that crew member's conduct.⁵⁶ Next, it may be difficult to determine who is a 'national', especially where an alleged offender holds dual nationalities or where one

⁵¹ NASA, 'Visitors to the Station by country' (*NASA*, 15 Nov 2020) <<https://www.nasa.gov/feature/visitors-to-the-station-by-country/>> accessed 23 February 2022.

⁵² Agreement among the United States of America, governments of Member States of the European Space Agency, the government of Japan, and the government of Canada on Cooperation in the Detailed Design, Development, Operation, and Utilisation of the permanently Manned Civil Space Station, (signed 29 September 1988, entered into force 30 January 1992) 1988, 37 *Zeitschrift für Luft-und Weltraumrecht* (ZLW) (hereinafter 'IGA 1988') s 22.

⁵³ *ibid* s 1.

⁵⁴ Ratner (n 3) 335.

⁵⁵ Ohmer (n 3) 371.

⁵⁶ Ohmer (n 3) 372.

Partner State asserts that an alleged offender is its national but other Partner States dispute that conclusion (see Section 2(b) below). This paper therefore argues that the ideal means of apportioning criminal jurisdiction in outer space is one that is not as overly reliant on nationality as the IGA is.

Additionally, Article XXII(2), under which the complainant nation may assert criminal jurisdiction if it does not believe the perpetrator's state can prosecute the perpetrator, suggests a deep mistrust among the Partners about each other's judicial systems.⁵⁷ There are already political difficulties with countries collaborating in outer space. Under US federal law, for example, NASA is precluded from partnering with China on space projects.⁵⁸ In response, China has opted to build its own lunar base instead of joining the ISS project. For its part, Russia's enmity with the US has led it to divest itself from its interest in the ISS to join China's lunar base project.⁵⁹ Basing a wider means of apportioning criminal jurisdiction in outer space on Article XXII(2)'s unilateralism would undoubtedly exacerbate those political tensions. The Outer Space Treaty, for example, which was negotiated among far more parties than the IGA, lacks Article XXII(2)'s default provision.

Finally, Article XXII(2) is also silent on Partner States exercising criminal jurisdiction over non-Partner States and non-Partner States exercising criminal jurisdiction over their own nationals. That would be difficult to apply to private spacecraft, where different people of different nationalities travel on spacecraft assembled by different nations. The IGA's simplicity is therefore something that can be learned from when apportioning criminal jurisdiction in outer space. But this wider means of apportionment must account for every signatory to the Outer Space Treaty — not just the IGA's Partner States — while avoiding one nation unilaterally asserting its criminal jurisdiction.

⁵⁷ Julian Hermida, 'Crimes in Space' (2006) 31 *Annals of Air & Space Law* 405, 411.

⁵⁸ Virginia A Seitz, 'Unconstitutional restrictions on activities of the Office of Science and Technology Policy in section 1340(A) of the Department of Defence and Full-Year Continuing Appropriations Act, 2011' (*US Department of Justice*, 19 September 2011) <<https://web.archive.org/web/20120713080223/http://www.justice.gov/olc/2011/conduct-diplomacy.pdf>> accessed 21 February 2022.

⁵⁹ Henry Foy, 'Russia to pull out of International Space Station in 2025' (*Financial Times*, 21 April 2021) <<https://www.ft.com/content/a1518565-e643-42ac-a650-02e9c3bdd657>> accessed 26 April 2022.

II. TRADITIONAL BASES OF PRESCRIBING CRIMINAL JURISDICTION ON EARTH

Because of the limits of the current means of apportioning criminal jurisdiction in outer space, international law relies on its traditional means of prescribing criminal jurisdiction on Earth to fill in the gaps. As with the law of the High Seas on Earth, the general starting point for outer space is that it is *res communis*: it belongs to everyone.⁶⁰ Furthermore, the Outer Space Treaty allows for international law on Earth to apply in outer space too (see Section 1(a) above). The Draft Convention on Jurisdiction with Respect to Crime — a comprehensive study of the principles for prescribing criminal jurisdiction — identified the four primary bases for doing so: territoriality, nationality, protectiveness, and universality.⁶¹ As this Section explains, however, each of these bases have flaws that prevent any of them from being the sole means of prescribing criminal jurisdiction in outer space.

a) Territorial principle

The territorial principle is the oldest means of apportioning criminal jurisdiction in international law, dating to the Treaty of Westphalia's idea that a nation's power ends at its borders.⁶² Under this approach, a nation has exclusive criminal jurisdiction within its borders. In the context in which a nation seeks to prescribe criminal jurisdiction, jurisdiction applies where either the alleged misconduct occurs in that state's territory (subjective territoriality), or where the effects of such alleged misconduct have a 'substantial effect within its territory'⁶³ (objective territoriality).

Territoriality would offer two benefits in outer space. First, it would be familiar to nations seeking to apportion criminal jurisdiction in outer space. Territoriality is not only the primary means of exercising criminal jurisdiction on

⁶⁰ Outer Space Treaty (n 6) art I.

⁶¹ Edwin D Dickinson and others, 'Draft Convention on Jurisdiction with Respect to Crime' (1935) 29 American Journal of International Law Supplement 439.

⁶² Leo Gross, 'The Peace of Westphalia, 1648 – 1948' (1948) 42(1) American Journal of International Law 20.

⁶³ Mary B McCord, 'Responding to the Space Station Agreement: the extension of US law into space' (1989) 77(5) Georgetown Law Journal 1933, 1939.

Earth⁶⁴ but, as per Article VII of the Outer Space Treaty, it is also the agreed-upon primary means of exercising criminal jurisdiction aboard spacecraft. Although nations may not assert sovereignty in outer space, something analogous to territoriality may still apply aboard spacecraft under international law's 'flag state jurisdiction', the analogy for which comes from ships sailing on the High Seas which, like outer space, is *res communis*.⁶⁵ Flag state jurisdiction treats a ship as the sovereign territory of the nation whose flag that ship flies.⁶⁶ Because Article VI of the Outer Space Treaty requires a nation to register any spacecraft or objects launched from that nation's jurisdiction,⁶⁷ flag state jurisdiction could apply to anything a state registers and launches into outer space. An agreed-upon method of apportioning criminal jurisdiction would not encounter the same political and legal difficulties that other, still novel, methods would (see Section 3(c) below). Flag state jurisdiction even has the advantage of acknowledging that people on sea ships and spaceships operate in similar environments — isolated for extended periods in a confined setting in a hostile environment.⁶⁸

Further emphasising this familiarity with flag state jurisdiction in outer space is the fact that a group of space law experts from Germany, Russia, and the US tested its viability through a Draft Convention on Manned Space Flight.⁶⁹ The Convention would grant criminal jurisdiction to the registrant nation, while creating a chain of command in which the spacecraft's commander would assume responsibility for the spacecraft and everyone on board. The commander would then be accountable to the Mission Director on Earth.⁷⁰ Such an explicit articulation of the chain of responsibility would provide legal certainty to the mission's participants. According to the Convention's drafters,

⁶⁴ *S.S. Lotus (France v Turkey)*, 1927 PCIJ Series A, No. 10 (September 7) 18-19.

⁶⁵ O'Keefe (n 4) 14.

⁶⁶ United Nations Convention of the Law of the Sea (adopted 10 December 1982, entered into force 16 November 1994) 1833 UNTS 397 (UNCLOS) art 94. See also Bin Cheng, *Studies in International Space Law* (Clarendon Press 1997), 387 and *Lauritzen v Larsen*, 345 U.S. 571 (1953) at 929, citing *United States v Flores*, 289 U.S. 137, 155-156.

⁶⁷ Outer Space Treaty (n 6) art VI.

⁶⁸ Ohmer (n 3) 366.

⁶⁹ Karl-Heinz Bockstiegel and others, 'Draft Convention on Manned Spaceflight' (*OPS-Alaska*, 13 December 2015)

<https://ops-alaska.com/IOSL/V7P1/1990_MannedSpaceFlightConvention_EN.pdf> accessed 27 March 2022 ('the Convention').

⁷⁰ *ibid* art IV.

the commander would be responsible for enforcing the registrant nation's criminal laws aboard the spacecraft.⁷¹ In addition to adopting the Outer Space Treaty's emphasis on international cooperation while using outer space, the Convention explicitly encourages such cooperation where individuals from multiple nations participate in these flights.⁷²

The Convention garnered support from the Board of Directors of the International Institute of Space Law, who recommended that Germany, Russia, and the US submit the Convention to the United Nations Committee on the Peaceful Uses of Outer Space's ('COPUOS') Legal Subcommittee. They did so in 1990. However, the COPUOS has failed to place the Convention's recognition on its agenda.⁷³

Second, while this paper is primarily concerned with apportioning the jurisdiction to prescribe criminal law, it would also be easiest for a nation in which a crime was committed to enforce a criminal law against the defendant in that scenario. That nation would have the easiest access to preserving evidence and compelling witnesses aboard a spacecraft to testify.⁷⁴

The territorial principle would, however, suffer from several flaws if it were the sole means of apportioning criminal jurisdiction in outer space. First, once outside a spacecraft's doors, territoriality would immediately conflict with Article II of the Outer Space Treaty, which prevents nations from erecting borders in outer space within which they could assert criminal jurisdiction under the territorial principle. Where, for example, someone commits an offence outside a spacecraft and immediately returns to the spacecraft, territoriality would not apply because a nation's territorial claim over the spacecraft via flag state jurisdiction would not extend to outer space. Territoriality would only occur once the perpetrator returns to the spacecraft after he or she commits the

⁷¹ *ibid.*

⁷² *ibid.*

⁷³ Nandasiri Jasentuliyana, *Perspectives on International Law* (1st edn, Martinus Nijhoff Publishers 1995) 465.

⁷⁴ United Kingdom Mission to the United Nations, 'United Kingdom of Great Britain and Northern Ireland: The Scope and Application of Universal Criminal Jurisdiction' (UN General Assembly, 23 March 2020) <https://www.un.org/en/ga/sixth/75/universal_jurisdiction/unitedkingdom_e.pdf> accessed 16 March 2022.

offence. The law is clear, however, that territoriality's jurisdictional nexus must exist at the time of the commission of the offence.⁷⁵ Although territoriality is the most prevalent means of prescribing criminal jurisdiction on Earth, this paper argues that Article II of the Outer Space Treaty should prevail where it conflicts with territoriality in outer space. The UN General Assembly specifically considered and rejected sovereignty-based territoriality claims outside spacecraft in outer space when negotiating the Outer Space Treaty. They did so because of Cold War-era risks of either the US or the Soviet Union asserting sovereignty over any part of outer space for warfare purposes.⁷⁶ As noted above, the international community was hesitant to depart from the Outer Space Treaty to accommodate the Bogota Convention's signatories' concerns because of the political upheaval that would entail. The international community should likewise be hesitant to depart from the Outer Space Treaty for similar political concerns, even for customary international law as applicable on Earth.

Next, there may be confusion about where a nation may exercise territorial jurisdiction because, unlike the 'High Seas' on Earth, there is no internationally accepted definition of what 'outer space' is.⁷⁷ Once outside of a spacecraft, it would therefore be difficult to determine when an individual is within a nation's airspace versus outer space. In the former case, that nation's jurisdiction would apply through territoriality. In the latter case, however, Article II of the Outer Space Treaty precludes a nation from enforcing territorial jurisdiction. The implications of this concern are slightly different to the first problem. Article II of the Outer Space Treaty deals with a jurisdictional dispute in an area of outer space in which it is internationally accepted that no nation can assert territorial jurisdiction. The definitional concern identified here, meanwhile, deals with a jurisdictional dispute in an area where at least one nation believes that it can rightfully assert territorial jurisdiction because there is a separate dispute about whether the incident arose in that nation's airspace or in an area of *res communis*, properly governed by Article II.

Additionally, as with the Outer Space Treaty, a dispute may arise over which nation ought to claim flag state jurisdiction for a spacecraft. Unlike most

⁷⁵ *S.S. Lotus* (n 64) 23.

⁷⁶ Stuart Banner, *Who Owns the Sky? The Struggle to Control Airspace from the Wright Brothers On* (Harvard University Press 2008) 284.

⁷⁷ Lotta Viikari, *The Environmental Element in Space Law* (Brill 2008) 1-2.

sea ships, where only one nation or company constructs the ship, many nations and companies may own parts of a full spacecraft, such as the ISS' modules that are owned by different Partner States. With spacecraft registration, however, only one nation may register a spacecraft.⁷⁸ This would ignore other nations' contributions to that spacecraft and raise the spectre of further political disputes arising.

Furthermore, nations would regard it as fundamentally unfair if nation B unilaterally asserted sole jurisdiction over acts a national of nation A commits in nation B. Correcting the territorial principle's unfairness dates to the 1927 *Lotus* case,⁷⁹ in which the Permanent Court of International Justice (PCIJ) allowed Turkey to assert criminal jurisdiction under Turkish national law against a French national. It was also what motivated the UN to allow signatories to the Registration Convention to negotiate exceptions to the general rule of criminal jurisdiction belonging to the country of registration.⁸⁰

Finally, a territorial means of prescribing criminal jurisdiction in outer space may create inequalities between nations, or corporations residing in particular nations, which can afford to build a spacecraft and those that cannot. For example, of the 4,852 active artificial satellites orbiting Earth as of January 1, 2022, 2,944 belonged to the US government or US-resident corporations.⁸¹ Recall that the IGA's Partner States negotiated the IGA 1998 to explicitly remove the US' unilateral criminal jurisdiction on the IGA from the IGA 1988. It is similarly likely that the other 194 sovereign nations on Earth may regard it as unfair if the US — whose artificial satellites represent over 56% of those orbiting Earth — could unilaterally prescribe criminal jurisdiction for over half of activity in outer space based on territoriality alone. Even the ISS, representing the greatest number of nations participating on one project in outer space, has only 16 Partner States. While the IGA's progress from the 1988 to the 1998 versions suggests that the Outer Space Treaty could similarly be amended as more nations register spacecraft, this does not solve the inequality issue. In a

⁷⁸ Registration Convention (n 19) Annex, art II (2) and (3).

⁷⁹ *S.S. Lotus* (n 64). See also Ireland-Piper and Freeland (n 3) 52-54.

⁸⁰ Lee Seshagiri (n 3) 483.

⁸¹ Erick Burgueño Salas, 'Number of satellites in orbit – major countries 2022' (*Statista*, 27 July 2022) <<https://www.statista.com/statistics/264472/number-of-satellites-in-orbit-by-operating-country/>> accessed 10 October 2022.

database compiled by Thomas G. Roberts as of September 1, 2022, for example, the cheapest small satellite — defined as a small-lift vehicle carrying up to 2,000 kg into low-Earth orbit — was the Shian Quxian, which cost approximately \$5 million USD to assemble and launch into outer space, a cost far beyond many countries' treasuries.⁸² Space activity as the domain of the wealthy would be exacerbated if criminal jurisdiction was based on territoriality. This is exacerbated by, as noted above, the US arguing against recognizing the Bogota Convention, an instrument that would have primarily benefited developing nations along the equator. Although the territorial principle is the primary means of prescribing criminal jurisdiction on Earth, the territorial principle would be politically and administratively unfeasible as being the sole means of prescribing criminal jurisdiction in outer space.

(b) Aboard aircraft

The idea of a 'floating flag state jurisdiction' would not be completely foreign in international law. Although not one of the four traditional means of prescribing criminal jurisdiction on Earth, the Tokyo Convention on Offences and Other Acts Committed on Board Aircraft 1963⁸³ ('Tokyo Convention') is a *sui generis* means of prescribing criminal jurisdiction aboard aircraft that enjoy international recognition. Article 3(1) of the Tokyo Convention accords the aircraft registrant's nation with criminal jurisdiction for incidents arising while the aircraft is in the air, on the surface of the High Seas, or anywhere else outside another nation's territory.⁸⁴ Much like flag state jurisdiction, aircraft jurisdiction is similar to territoriality, in which the 'territory' is assumed to be the vessel, whose identity is the nation in which the vessel is registered.

Although aircraft jurisdiction under the Tokyo Convention would face similar challenges in outer space as territoriality, it may solve at least one of territoriality's problems. Were a spacecraft subject to the Tokyo Convention's

⁸² Thomas G Roberts, 'Space launch to low Earth orbit: how much does it cost?' (*Aerospace Security*, 27 June 2022) <<https://aerospace.csis.org/data/space-launch-to-low-earth-orbit-how-much-does-it-cost/>> accessed 10 October 2022.

⁸³ Convention on offences and certain other acts committed on board aircraft (adopted 14 September 1963, entered into force 4 December 1969) 220 UNTS 136 (Tokyo Convention).

⁸⁴ Termed 'aircraft jurisdiction' for ease of reference.

aircraft jurisdiction, it would not matter where the boundary lies between where a nation's airspace ends and outer space begins, at least for incidents occurring on board the spacecraft. In those situations, the nation in which the spacecraft is registered would assume jurisdiction.

Aircraft jurisdiction, however, would face a similar challenge as flag state jurisdiction in outer space. Namely, aircraft jurisdiction has no answer for incidents arising outside of an aircraft — or spacecraft, as it were.

c) Nationality principle:

Nationality may be divided between active nationality and passive personality. Active nationality provides jurisdiction to a nation over the acts of that nation's nationals, regardless of those nationals' locations. The IGA's Article XXII(1) is one example of active nationality. In contrast, passive personality grants jurisdiction to a nation where a victim is a national of that nation. It is typically done so for serious offences against the person, such as homicides or sexual offences, because of the perceived fairness of allowing a victim's nation the opportunity to try such a case.⁸⁵ Although controversial in the past,⁸⁶ passive personality is generally recognized as customary international law today.⁸⁷

Traditionally, nationality has three ideas inherent to it: allegiance,⁸⁸ membership of a political community,⁸⁹ and 'social facts of attachment'.⁹⁰ Here, nationality may be distinguished from citizenship. The latter is a domestic law concept, typically determined under more formalistic legal rules.⁹¹ The former,

⁸⁵ *Case Concerning the Arrest Warrant of 11 April 2000 (Democratic Republic of the Congo v Belgium)* [2002] ICJ Rep 1 (14 February 2002) ('*Arrest Warrants*') 46-47, 51-52 (Joint Separate Opinion of Judges Higgins, Koojimans, and Buergenthal).

⁸⁶ *ibid* [47] (Joint Separate Opinion of Judges Higgins, Koojimans, and Buergenthal).

⁸⁷ See, for example, the Rome Statute of the International Criminal Court (adopted 17 July 1998, entered into force 1 July 2002) 2187 UNTS 90 (Rome Statute) art 90(6)(b), UNSC Res 1816 (2 June 2008) UN Doc S/RES/1816 11, UNSC Res 1846 (2 December 2008) UN Doc S/RES/1846 14, and United Nations Transitional Administration in East Timor Regulation No. 2000/15 s. 2.2(c).

⁸⁸ Rishi Gulati, "Resolving dual and multiple nationality issues in a globalised world" (2014) 28(1) *Journal of Immigration Asylum and Nationality Law* 27, 28.

⁸⁹ *James Hammet Howard (Great Britain) v United Mexican States* (1931) 5 RIAA 232-233.

⁹⁰ *Liechtenstein v Guatemala (Second Phase)* [1955] ICJ Rep 4 ('*Nottebohm Case*') 23.

⁹¹ Gulati (n 88) 35.

meanwhile, is an international law concept, with more fluidity in determining how relevant each of the three ideas inherent to it are.

Nationality would offer three benefits in outer space. First, adopting nationality as a means of prescribing criminal jurisdiction in outer space would avoid territoriality's political challenges. Not only does it allow for the recognition of both the perpetrator and victim's state's jurisdiction, but it also implicitly acknowledges that professional astronauts may be representatives of their national states to whom sovereign immunity or international courtesy may apply.⁹² Second, it also implicitly acknowledges the practical realities astronauts face when living in spacecraft. Rather than confining themselves to their home nation's modules, within which territoriality would apply, astronauts work across the entire spacecraft or space station, which under the territoriality principle may be in other nations' jurisdictions.⁹³ Third, nations may be familiar with using nationality in outer space because it is already the primary means of apportioning criminal jurisdiction in Antarctica, another 'admittedly sovereignless land',⁹⁴ and under the IGA. Canada has also proposed an amendment to its domestic Criminal Code to establish nationality-based criminal jurisdiction over its nationals on the Moon, the Lunar Gateway (a man-made facility orbiting the Moon), or while in transit between Canada, the Moon, and the Lunar Gateway.⁹⁵ The proposed amendment is as follows:

A Canadian crew member who, during a space flight, commits an act or omission outside Canada that if committed in Canada would constitute an indictable offence is deemed to have committed that act or omission in Canada.⁹⁶

Nationality would, however, suffer from three drawbacks in outer space. First, it would deny a nation jurisdiction over non-nationals even where

⁹² Chatzipanagiotis (n 36) 112.

⁹³ *ibid.*

⁹⁴ *Beattie v United States*, 756 F.2d 91, 94 (D.C. Cir. 1984). See also the Antarctic Treaty (adopted 1 December 1959, entered into force 23 June 1961) 402 UNTS 71 (Antarctic Treaty) art VIII.

⁹⁵ Bill C-19, An act to implement certain provisions of the budget tabled in Parliament on April 7, 2022 and other measures, 1st Session, 44th Parliament, 2022.

⁹⁶ *ibid* s. 296(4)(2.35).

that nation may have a legitimate interest in prosecuting that non-national under its domestic law. For example, although all NASA astronauts travelled to outer space in Russian-registered Soyuz rockets between 2011 and 2020,⁹⁷ a nationality-based means of prescribing criminal jurisdiction would leave Russia with little authority aboard its spacecraft unless the alleged offenders were Russian. Such a denial of jurisdiction may exacerbate political tensions in outer space if an alleged offender were subject to prosecution between multiple nations' penal codes, particularly between a nation claiming territorial jurisdiction (Russia, for example, on its Soyuz spacecraft) versus another claiming nationality jurisdiction. As of right now, there are no means of determining which jurisdiction ought to prevail in outer space, although some commentators have argued for such a forum.⁹⁸ This concern is evident with proposed Canadian legislation, which would grant Canada jurisdiction in outer space over its nationals, regardless of whether those nationals were on a vessel registered in another country or not.

The Canadian legislation's application to 'indictable offences' presents another problem. Unlike in English law, a quirk in Canadian domestic law allows the Crown to decide for 'hybrid offences' (sexual assault, for example) whether to proceed by indictment or summarily. The Crown normally bases its decision on the seriousness of the accused's alleged actions and the harm caused.⁹⁹ With this proposed legislation, however, there is a risk that the Attorney General — as a member of Cabinet and the government's chief prosecutor — may exert political influence over the Crown to proceed via indictment to claim jurisdiction over an alleged offence in outer space. In so doing, the risk for other countries is that Canada may seek to artificially assert a claim for jurisdiction by electing to proceed via indictment. This is especially relevant where passive personality is concerned. As noted above, countries typically invoke passive personality where there is a serious offence against a person, such as a sexual offence. If a Crown Attorney would have normally deemed it in the public

⁹⁷ Daniel Oberhaus, 'The US Hitches Its Final Ride to Space From Russia – for Now' (*Wired*, 08 April 2020) <<https://www.wired.com/story/the-us-hitches-its-final-ride-to-space-from-russia-for-now/>> accessed 20 October 2022.

⁹⁸ See, for example, Caleb Ohmer (n 3), Lee Seshagiri (n 3), and Stacey Ratner (n 3) 341.

⁹⁹ Government of Canada, 'Criminal Offences' (*Government of Canada*, 7 July 2021) <<https://www.justice.gc.ca/eng/cj-jp/victims-victimes/court-tribunaux/offences-infract.html>> accessed 2 July 2022.

interest to pursue the sexual offence charge summarily, the risk is that the Attorney General may seek to ‘upgrade’ the charge to an indictable offence to give Canada potential jurisdiction over the crime.¹⁰⁰ Parliament was alive to such concerns in other parts of the Criminal Code, such as the requirement for the Attorney General to specifically not consider political influence when deciding whether to grant a deferred prosecution agreement.¹⁰¹ But there is no such saving provision for the proposed outer space jurisdiction legislation.

Nationality’s second problem is that it may be difficult to determine. A nation may regard a person to be its national but other nations may disregard that status for jurisdictional purposes. In *Nottebohm’s Case*, for example — the leading case in resolving disputes over nationality — Mr. Nottebohm had German citizenship by birth, though spent most of his life in Guatemala, eventually obtaining permanent residency there. Liechtenstein later recognized Mr. Nottebohm as one of its citizens when Mr. Nottebohm applied for and obtained Liechtenstein citizenship in 1939. Guatemala, however, disregarded Mr. Nottebohm’s Liechtenstein citizenship, believing it to be a sham to evade Guatemala’s domestic ‘enemy alien’ law — which targeted German citizens — drafted in response to World War II. The Guatemalan authorities subsequently arrested Mr. Nottebohm in 1943. The International Court of Justice held that Liechtenstein lacked standing to bring a claim on Mr. Nottebohm’s behalf because Mr. Nottebohm’s ‘effective nationality’ was Guatemalan, not Liechtensteiner. The Court based its decision on the factual ties Mr. Nottebohm maintained with Guatemala as being ‘the main seat of his interests’.¹⁰²

Nationality may also be difficult to determine because a person’s ‘choice’ is not a traditional means of prescriptive criminal jurisdiction in international criminal law.¹⁰³ Dual citizens would potentially face both problems. For private corporations operating in outer space, although they too would have nationality, the rules for determining that nationality are complex and subject to abuses, such as forum shopping.¹⁰⁴ One way to resolve any uncertainty over nationality would be to issue ‘space visas’, for which the nation who issues the

¹⁰⁰ *Criminal Code*, RSC 1985, c C-46, s 601.

¹⁰¹ *ibid* s. 715.32(3).

¹⁰² *Nottebohm Case* (n 90) 22. See also Section 5(b) below.

¹⁰³ *ibid* [43].

¹⁰⁴ *ibid* [71].

visa would assume criminal jurisdiction.¹⁰⁵ This would be analogous to travellers presenting their visas at the airport to enter a country. The clear appeal of a space visa regime is that it would be simple to implement because of its similarity to travel visas. The main problem with a space visa regime is that, like the Registration Convention, it only grants primary jurisdiction *prima facie*. It does not resolve a dispute where there are competing claims. They also fail to address who would assume criminal jurisdiction for transnational organised crimes, such as migrant smuggling, environmental contamination, or weapons trafficking, for which there may be people involved from several visa-granting nations.¹⁰⁶

Nationality's third problem is that, although the Antarctic Treaty's emphasis on peace and collaboration mirrors the Outer Space Treaty,¹⁰⁷ the former's emphasis on nationality has similar weaknesses to the IGA's. Namely, both the Antarctic Treaty and the IGA's nationality principle only apply to some individuals. With the Antarctic Treaty only observers under the treaty, scientists, and staff are subject to national jurisdiction. However, for tourists, military personnel, and non-privileged foreign nationals, the Antarctic Treaty's relevant signatories are only supposed to 'consult together with a view to reaching a mutually acceptable solution'.¹⁰⁸ As with the IGA and the Antarctic Treaty, nations may be unwilling to consult with one another to resolve nationality-based disputes — and thus disputes over jurisdiction — for fear of appearing to cede a claim of sovereignty.¹⁰⁹

d) Protective principle:

The protective principle allows a nation to assert criminal jurisdiction over specific acts, regardless of who committed the acts and where they did so, that might affect that nation's security or interests.¹¹⁰ On Earth, nations have used the protective principle to assert criminal jurisdiction for crimes such as

¹⁰⁵ Blount (n 3) 301.

¹⁰⁶ Ireland-Piper and Freeland (n 3) 70.

¹⁰⁷ Antarctic Treaty (n 94) Preamble.

¹⁰⁸ *ibid* art VIII.

¹⁰⁹ Todd Chatham, 'Criminal jurisdiction in Antarctica: a proposal for dealing with jurisdictional uncertainty and lack of effective enforcement' (2010) 24 *Emory International Law Review* 331, 337.

¹¹⁰ Ohmer (n 3) 364.

‘counterfeiting currency, desecration of flags, economic crimes, forgery of official documents such as passports and visas, and political offences such as treason’.¹¹¹ However, because of the threats to relations between nations by invoking the protective principle, many national penal codes restrict its use.¹¹² In one case, for example, the US prosecuted an East German citizen under the US’ Espionage Act for selling classified intelligence to foreign operatives in Mexico and East Germany. In that case, however, the district judge specifically held that protectivity could apply because the nature of the crime (espionage) did not depend on locality.¹¹³ The judge also held that espionage was a serious enough crime against the State’s functioning to warrant interpreting Congress’ intent to punish as giving the statute extraterritorial effect.¹¹⁴ Most crimes, however, require locality and while those crimes may be serious, they do not represent a direct threat against a nation’s functioning. And while the assault may be a life-altering matter to the victim, it is not a direct threat to a nation’s functioning. The use of the protective principle should likewise be limited in outer space for the same reason.

One area in which the criminal jurisdiction based on protectivity may have use is with terrorism, as international law defines it to include ‘actions that undermine civilians’ essential rights (namely universal values, such as life, physical integrity, freedom, and dignity) in a manner likely to receive absolute condemnation by the whole international community’.¹¹⁵ One concern, however, may be determining who ought to have prescriptive criminal jurisdiction based on protectivity where a terrorist incident affects multiple nations.

e) Universal principle:

Universal criminal jurisdiction is ‘jurisdiction over offences committed extraterritorially by non-nationals against non-nationals, where the offence constitutes no threat to the fundamental interests of the prescribing [nation] and

¹¹¹ Danielle Ireland-Piper, *Accountability in Extraterritoriality: A Comparative and International Law Perspective* (Edward Elgar Publishing Limited 2017) 33.

¹¹² Ohmer (n 3) 364.

¹¹³ *United States v Zebe*, 601 F. Supp. 196 (D. Mass. 1985) 200.

¹¹⁴ *ibid.*

¹¹⁵ Marcello Di Filippo, ‘The definition(s) of terrorism in international law’ in Ben Saul (ed) *Research handbook on international law and terrorism* (Edward Elgar Publishing 2020) 15; O’Keefe (n 4) 12.

does not give rise to effects within its territory'.¹¹⁶ Countries exercise universal jurisdiction where the other bases of criminal jurisdiction would not apply, such as a war crime that did not occur in that nation's territory. In such a case, the prescribing nation exercises jurisdiction not because it poses a threat to that nation's fundamental interests, but because the alleged crime is serious enough for either that prescribing nation to act as a 'global enforcer' of the law or that the accused should not be offered a 'safe haven' from the law.¹¹⁷ Although two of the judges in *Case Concerning the Arrest Warrant of 11 April 2000 (Democratic Republic of the Congo v Belgium)* ('*Arrest Warrants*') described 'universal jurisdiction *in absentia*' as problematic,¹¹⁸ universal jurisdiction is commonly accepted for war crimes, crimes against humanity, and genocide, as defined by customary international law.¹¹⁹ Should any of those acts occur in outer space, even in a military context, it is doubtful that nations would object to universal jurisdiction being exercised.¹²⁰

More controversy would arise from prescribing universal jurisdiction over piracy *jure gentium* ('as per the law of nations') in outer space because its potentially wide ambit directly conflicts with the Outer Space Treaty's Article VIII flag state jurisdiction. As per Article 101 of the United Nations Convention on the Law of the Sea ('UNCLOS'), 'piracy' consists of, among other things, illegal acts of violence or detention committed for private ends by crew or passengers of a private ship or aircraft on the high seas.¹²¹ International law considers both outer space and the High Seas to be *res communis*. It would therefore be relatively simple at first glance to transplant a universal criminal jurisdiction to outer space for acts of piracy. 'Private ends' refers to 'non-state actors ... without authorization by public authority'.¹²² 'Illegal' is a reference to domestic law, meaning it lies with nations to define what counts as 'violence' or

¹¹⁶ O'Keefe (n 4) 17.

¹¹⁷ *ibid.*

¹¹⁸ *Arrest Warrants* (n 85) 12, 15-16 (Separate Opinion of President Guillaume), 9-11 (Declaration of Judge Ranjeva).

¹¹⁹ *ibid* [173].

¹²⁰ Icho Kealotswe-Maltou, 'The rule of law in outer space: a call for an International Outer Space Authority' in Cassandra Steer and Matthew Hirsch (eds), *War and Peace in Outer Space: Law, Policy, and Ethics* (Oxford University Press 2021) 105.

¹²¹ UNCLOS (n 66) art 101.

¹²² *Republic of Seychelles v Ahmed and Five Others*, Crim Side No 21 of 2011, 14 July 2011, para. 21.

‘detention’.¹²³ Transplanting ‘piracy’s’ terrestrial definition to outer space would therefore clearly cover many acts in the latter.

The international community should, however, restrict piracy *jure gentium*’s ambit in outer space because, rather than supplementing existing practice in outer space, which this paper argues for, a universal jurisdiction for piracy *jure gentium* would contradict it. Applying Article 101 of UNCLOS to outer space would leave universal jurisdiction for piracy *jure gentium*’s ambit far broader than on Earth. The High Seas comprise over 50% of Earth’s surface area,¹²⁴ but 100% of outer space is *res communis*. Every single act of violence or detention — as defined domestically — for non-state purposes aboard a private spacecraft could therefore be subject to universal jurisdiction. To allow such widespread use of universal jurisdiction would be contrary to the Outer Space Treaty’s Article VIII’s spirit, which allows for flag state jurisdiction of such acts aboard registered spacecraft, whether private or publicly owned. One solution to this may be for the nation that could have exercised flag state jurisdiction to grant authority to another nation seeking to exercise universal jurisdiction for piracy *jure gentium*, at the former’s discretion. This would be more in line with the spirit of Article II(2) of the Registration Convention, which allows nations to negotiate exceptions to Article VIII of the Outer Space Treaty. One concern with this proposition is that, as seen in the *S.S. Lotus* case, countries are naturally reluctant to cede their jurisdiction to prescribe criminal law.

III. HIERARCHICAL CLAIMS

(a) On Earth

It is generally accepted that there is no hierarchy among concurrent prescriptive criminal jurisdictions on Earth because of the political and legal difficulties of deciding which means of prescribing jurisdiction should apply in each case. In the *SS Lotus* case, for example, the Court held that, because an individual committed a criminal act on a French ship, whose victims were on a Turkish ship, Turkey could try the individual on the French ship in Turkey’s

¹²³ *ibid.*

¹²⁴ Richard Blaustein, ‘United Nations Seeks to Protect High-Seas Biodiversity’ (2016) 66(9) *BioScience* 713.

criminal courts because there was no rule in international law that prevented Turkey from doing so. According to the Court:

The offence ... was an act ... having its origins on board the [French ship], whilst its effects made themselves felt on board the [Turkish ship] ... These two elements are, legally, entirely inseparable ... Neither the exclusive jurisdiction of either State, nor the limitations of the jurisdiction of each to the occurrences which took place on the respective ships would appear calculated to satisfy the requirements of justice and effectively to protect the interests of the two States ... It is therefore a case of concurrent jurisdiction.¹²⁵

The Draft Convention on Jurisdiction with Respect to Crime specifically rejected such a hierarchy as ‘unwarranted by anything in international law and unsupported by the existing practice of [nations]’.¹²⁶ Likewise, modern multilateral treaties reject such a hierarchy of prescriptive criminal jurisdictions where there are concurrent jurisdictions,¹²⁷ instead requiring nations claiming concurrent criminal jurisdiction to consult with one another.¹²⁸

This practice on Earth should not, however, carry over to outer space for two reasons. First, a hierarchy in outer space would fill the gaps in the existing law identified in Sections 1 and 2 without violating existing international law. As Section 1 above argued, Article VIII of the Outer Space Treaty clearly places territoriality — or at least its flag state jurisdiction equivalent — at the fore, with nationality applying aboard the ISS via Article XXII(1) of the IGA. However, as Sections 1 and 2 also highlighted, there are enough gaps within the Outer Space Treaty and territoriality that they could not be the only means of

¹²⁵ *S.S. Lotus* (n 64) 30, 31.

¹²⁶ Dickinson and others (n 61) 583.

¹²⁷ See, for example, Ronald Schmidt, ‘Convention Against Torture and Other Cruel, Inhuman, or Degrading Treatment or Punishment, Part 1 Substantive Articles, Art. 5 types of jurisdiction over the offence of torture’ in Manfred Nowak and others (eds) *The United Nations Convention Against Torture: A Commentary* (Oxford University Press 2019) 210. Mr. Schmidt argues that the drafters of the Convention Against Torture and Other Cruel, Inhuman, or Degrading Treatment or Punishment 1984 specifically rejected applying a hierarchy among prescriptive jurisdictions in the Convention for torture.

¹²⁸ See the treaties cited in O’Keefe (n 4); (n 116).

prescribing criminal jurisdiction in outer space. Something else, even if secondary, would be needed to fill the gaps territoriality leaves behind.

Second, there would be no risk of a hierarchy in outer space potentially disrupting the ‘existing practice of nations’ on Earth. The gaps identified at Sections 1 and 2 above (among others ‘outer space’ being undefined, spacecraft being assembled by multiple nations, and spacecraft being unregistered) are specific to outer space and thus have an isolated effect.

(b) In outer space

As explained in Sections 1 and 2, there is no single general principle in international law that can account for many common scenarios under which it would be necessary to prescribe criminal jurisdiction in outer space. Inserting a hierarchy of criminal jurisdictions into Article VIII of the Outer Space Treaty would most effectively account for all of them. The hierarchy would work as follows: the first two principles (territoriality and active nationality) should be distinguished based on which is most practical to adopt in each situation. The next three principles (passive personality, universality, and protectiveness) should be distinguished based on the type of crime. Where a higher ranked jurisdiction covers a case and the nation entitled to claim jurisdiction has not waived it, that nation should be entitled to claim jurisdiction. However, where a particular jurisdictional base fails to address a situation, or where a nation entitled to claim jurisdiction declines to do so, the nation entitled to claim jurisdiction the next level down should have the opportunity to do so. This would continue all the way down the hierarchy.

The territoriality principle, or at least its flag state and aircraft jurisdiction analogies — representing the nation in which a spacecraft is registered — should be at the top of a hierarchy of prescribing criminal jurisdiction in outer space. Criminal jurisdiction would be based on the nation of the spacecraft’s registration. The Outer Space Treaty’s territorial principle enshrined in Article VIII has strong precedential value, both for its wide application on Earth and its acceptance in outer space.¹²⁹ Not only would a first-ranked territorial principle align with the political consensus reached by the

¹²⁹ Cestmir Cepelka and Jamie H C Gilmour, ‘The application of general international law in outer space’ (1970) 36(1) *Journal of Air Law and Commerce* 30, 34-35.

United Nations in finding that it best respects the ‘humanity and sovereignty of mankind’, but several countries have also used the Outer Space Treaty as a means of prescribing criminal jurisdiction in outer space in their own domestic laws.¹³⁰

Where the default jurisdiction of territoriality does not apply, active nationality should rank next as a means of prescribing criminal jurisdiction in outer space. As argued above, while territoriality is the dominant means of prescribing criminal jurisdiction on Earth,¹³¹ the international community has accepted nationality as partially applying in Antarctica.¹³² Antarctica, much like outer space, is an area over which no nation may unilaterally assert territorial sovereignty.¹³³ Unlike outer space, there is no possibility of a nation invoking flag state jurisdiction over Antarctica, as the Antarctic Treaty specifically precludes nations from asserting ‘new claim[s]’ or ‘enlargement of ... existing claim[s]’.¹³⁴ Two scenarios in outer space where nationality might apply would be where an alleged crime occurred outside of a spacecraft or on board a spacecraft that multiple nations assembled and there is a dispute about which nation that spacecraft ought to be registered in. Criminal jurisdiction would be based on the alleged offender’s nationality, which in most cases would be the same as his or her citizenship.

For collisions, it would be appropriate to apply similar rules as those applicable to the high seas. Under UNCLOS article 97, for example, criminal jurisdiction lies with either the accused’s ship’s flag state or the accused’s nationality.¹³⁵ For practical purposes, however, it is usually the flag state who exercises jurisdiction because, per UNCLOS article 97(3), only the flag state can detain a vessel for investigative purposes.¹³⁶ Additionally, UNCLOS Article

¹³⁰ See, for example, the United Kingdom’s Outer Space Act 1986, c 38, s 12; Russia’s Law of the Russian Federation ‘About Space Activity’ Decree 5663-1 of the Russian House of Soviets 1993, art 9(3); Russia’s Statute No. 104 – Statute on Licensing Space Operations 1996, ss 2, 3.

¹³¹ *S.S. Lotus* (n 64) 16-17.

¹³² Antarctic Treaty (n 94) art VIII (1).

¹³³ *ibid* art IV (2).

¹³⁴ *ibid*.

¹³⁵ UNCLOS (n 66) art 97.

¹³⁶ Changwoo Ha, ‘Criminal jurisdiction for ship collision and marine pollution in high seas – focused on the 2015 judgement on M/V Ernest Hemingway case’ (2020) 4(1) *Journal of International Maritime Safety, Environmental Affairs, and Shipping* 11.

94(7)'s requirement for the flag state to 'cause an inquiry' into incidents involving the flag state ship that cause 'loss of life or serious injury ... or serious damage to ships or installations ... or to the marine environment' suggests the international community's intention to grant primary criminal jurisdiction to the flag state in such cases.¹³⁷

Where, however, the alleged offender is a dual national, or another nation disputes the nationality, nationality should be based on a factual determination of the legal relationship between the alleged offender and the nation asserting jurisdiction under active nationality.¹³⁸ Such facts may include long standing residence in the claimant nation, the degree to which the person has close personal or family ties to anyone in the claimant nation, whether the person demonstrated an intention to remain in the claimant nation, whether there was an absence of close family or personal ties in the other claimant nation, and the person's choice of nation.¹³⁹

Active nationality would address common situations that the territorial principle either ignores or fails to fully engage with, such as incidents arising outside of a spacecraft, and incidents arising in spacecraft assembled by multiple nations, none of whom would want to surrender criminal jurisdiction based on registration. Situations such as the ISS illustrate this point. As noted above, the IGA 1998 was specifically negotiated to preclude the possibility of one nation exercising exclusive jurisdiction over an entire spacecraft.¹⁴⁰ There may, however, be situations where it would be impossible to completely abandon a mission, despite political challenges on Earth, but that nation would still seek to exercise some degree of jurisdiction over its nationals. On the ISS, for example, there are four US nationals and five Russian nationals. Despite heightened tensions between the US and Russia over the war in Ukraine, Russia has stated that it has delayed abandoning the ISS until 2028 to permit its current astronauts to remain on the ISS to complete their ongoing missions.¹⁴¹ Active nationality would also

¹³⁷ *ibid.*

¹³⁸ Gulati (n 88) 35.

¹³⁹ *ibid.*

¹⁴⁰ See, for example, Ratner (n 3) 335 and Ohmer (n 3) 371.

¹⁴¹ Joey Roulette, 'Russia tells NASA space station pullout less imminent than indicated earlier' (*Reuters*, 27 July 2022)

,<https://www.reuters.com/business/aerospace-defense/russia-nasa-sticking-with-space-station-until-least-2028-2022-07-27/> accessed 21 September 2022.

resolve any inequalities that territoriality creates, since criminal jurisdiction would not be restricted to only those nations who could afford to send spacecraft into outer space. It would also have some precedential value in outer space — albeit less than territoriality — since it would accord with the IGA.¹⁴² Although the IGA has fewer signatories than the Outer Space Treaty and only applies to the ISS, nationality would still be a means of apportioning criminal jurisdiction in outer space to which many space-going countries are accustomed.

Beneath active nationality, passive personality should rank third as a means of prescribing criminal jurisdiction in outer space. Criminal jurisdiction would be based on the victim's nationality, which, like active nationality, would be the same as citizenship. However, where active nationality could apply to a wide array of crimes, nations should restrict exercising criminal jurisdiction in outer space based on passive personality to serious offences against the person, such as homicides or sexual offences. This would accord with the increasingly customary practice on Earth,¹⁴³ facilitating its adoption into outer space.

Deciding between universality and protectiveness for fourth depends on whether the act in question affects a single nation's vital interests or the interests of the international community.¹⁴⁴ In the former case, protectiveness should prevail. Criminal jurisdiction would be based on whether the alleged offence affected the nation seeking prescriptive jurisdiction's 'fundamental interests or security', as determined by international law. Examples may include counterfeiting currency, desecration of flags, economic crimes, forgery of official documents such as passports and visas, political offences such as treason, and terrorism.

Where the act in question affects the interests of the international community, however, universality should rank fourth as a means of prescribing criminal jurisdiction in outer space. Criminal jurisdiction would be based on whether the alleged offence in question violated customary international law, such as war crimes, crimes against humanity, slavery, or genocide. It may also

¹⁴² IGA 1998 (n 7) art V(2).

¹⁴³ Gillian Doreen Triggs, *International Law: Contemporary Principles and Practices* (LexisNexis Butterworths 2006) 356.

¹⁴⁴ Yoram Dinstein, 'The universality principle and war crimes' in Michael N Schmitt and Leslie C Green (eds) *The Law of Armed Conflict: Into the Next Millennium* (US Naval War College 1998) 22.

cover piracy *jure gentium*, subject to the nation with flag state jurisdiction granting the nation seeking universal prescriptive jurisdiction the authority to exercise it. This would both replicate existing practice on Earth,¹⁴⁵ while respecting the Outer Space Treaty's Article VIII and complying with the spirit of the Registration Convention's Article II(2).

One challenge that this proposed hierarchy would face is if the hierarchy reached its final stage and several countries claimed universal jurisdiction or sought to deny a country from claiming jurisdiction under protectiveness. For example, an individual or group may detonate a kinetic weapon in orbit, potentially injuring a group of space-going individuals from one nation. Despite the attack being focused on one nation's individuals, several nations would regard the detonation of a kinetic weapon in Earth's orbit as affecting those nations' 'fundamental interests or security' because it would affect all of humanity's space activities. At this point, it would be a political question as to which nation ought to have prescriptive jurisdiction: protectivity for the nation whose citizens were directly injured or universality for what may be regarded as a crime against humanity.¹⁴⁶

(c) A 'minimum contacts model' alternative

Karen Robbins' argument for a 'minimum contacts model' to apply to outer space deserves special mention as, at first glance, it bears close resemblance to a hierarchy of criminal jurisdiction.¹⁴⁷ The minimum contacts model — a product of US civil law's due process provision — prescribes jurisdiction based on the accused's nexus to the nation asserting jurisdiction, and the degree to which the nation asserting jurisdiction has an interest in prosecuting the offence.¹⁴⁸ This would be based on a series of factors weighed similarly to a hierarchy for which this paper advocates, such as the accused's nationality, his or her centre of interests, the place of contract (if any), the

¹⁴⁵ See UNCLOS (n 66) art 101 and *Republic of Seychelles v Ahmed and Five Others* (n 122).

¹⁴⁶ Jinyuan Su, 'The legal challenge of arms control in space' in Cassandra Steer and Matthew Hersch (eds), *War and Peace in Outer Space* (Oxford University Press 2021) 181-185.

¹⁴⁷ Robbins (n 3).

¹⁴⁸ *International Shoe Co v State of Washington Office of Unemployment Compensation and Placement*, 326 U.S. 310 (1945) 316.

inaccessibility of a foreign forum, and the policy of the forum seeking jurisdiction.¹⁴⁹

A minimum contacts model would, however, be problematic in outer space in ways that hierarchical jurisdiction would not be. With respect, Ms. Robbins fails to engage with the difficulties with applying general principles of domestic law to fill gaps in due process in international law. While some scholars argue such gap-filling ought to be possible in some areas of international law in the context of due process,¹⁵⁰ these arguments do not fully engage with the fundamental disagreement between nations on the role of general principles, such as due process, in international law.¹⁵¹ These problems are not evident under a hierarchy of existing international law principles. As noted above, nations generally acknowledge the existence of territoriality, active personality, passive personality, universality, and protectiveness (though this is subject to some dispute). The main dispute here is resolving which principle applies in each circumstance. This paper's hierarchy aims to solve that concern.

Ms. Robbins and her successors also fail to acknowledge that, although Article 38(1)(c) of the Statute of the International Court of Justice requires the International Court of Justice ('ICJ') to apply 'the general principles of law recognized by civilized nations',¹⁵² the ICJ rarely does so. In a comprehensive survey of 126 contentious cases before the ICJ, for example, none were decided on a general principle of law.¹⁵³ With due process as interpreted under US law specifically, it is doubtful whether it would have sufficiently universal application

¹⁴⁹ *Lauritzen v Larsen*, 345 U.S. 571 (1952).

¹⁵⁰ Charles Kotuby Jr. and Luke Sobota, *General Principles of Law and International Due Process: Principles and Norms Applicable in Transnational Disputes* (Oxford University Press 2021).

¹⁵¹ Martins Paparinskis, 'Book reviews: General principles of law and international due process: principles and norms applicable in transnational disputes' (2019) 30(2) EJIL 689, 691.

¹⁵² TS No 993, art 38(1)(c).

¹⁵³ The International Law Association's Study Group on the Use of Domestic Law Principles, 'The use of domestic principles in the development of international law – working session report Sydney 2018' (*International Law Association*, 2018) <https://www.ila-hq.org/en_GB/documents/conference-report-sydney-conference-2018> accessed 20 October 2022.

for other countries to agree to it as a rationale for prescribing criminal jurisdiction in outer space.¹⁵⁴

With international criminal law specifically, an overreliance on general principles such as due process may potentially violate the principle of legality, especially the requirement of notice to the accused — similar to a nationality-based means of jurisdiction — and the strict construction of statutes.¹⁵⁵

CONCLUSION

Anne McClain is one of the few people in history to see Earth from above. She is also the only person as of this paper's submission to potentially fall foul of criminal law in outer space. She will not be the last. As more people enter outer space, Hans Sinha's question will become more poignant: 'Human nature being what it is ... what criminal law will guide and judge the behaviour of mankind in space?'¹⁵⁶ This paper argues that a hierarchy of prescriptive criminal jurisdictions would be the best answer to Dr. Sinha's question.

The traditional means of prescribing criminal jurisdiction in outer space — the Outer Space Treaty and the IGA — are insufficient as constructed. They were drafted in a time when only state actors ventured into outer space, for political competition and potential warfare. They do not account for the modern commercial purposes of using outer space. The four traditional means of prescribing criminal jurisdiction on Earth, meanwhile — the territoriality, nationality, universality, and protective principles — are also each unsuitable alone in outer space. They were designed for criminal incidents arising on Earth, and thus fail to account for common scenarios in outer space. Among the most important and potentially common of those scenarios include incidents occurring outside of spacecraft, on or involving unregistered spacecraft, on or involving a spacecraft that many nations assembled. They also include incidents arising between people of different nationalities on spacecraft.

¹⁵⁴ *ibid.*

¹⁵⁵ Neha Jain, 'Judicial lawmaking and general principles of law in international criminal law' (2016) 57 *Harvard International Law Journal* 111, 116.

¹⁵⁶ Hans P Sinha, 'Criminal Jurisdiction on the International Space Station' (2004) 30(1) *Journal of Space Law* 85, 86.

A better way of prescribing criminal jurisdiction in outer space would be to establish a hierarchy of jurisdictions, in which territoriality is first, active nationality is second, passive personality is third, universality is fourth, and protectivity is fifth. Although international law on Earth has consistently held that hierarchical criminal jurisdiction should not exist, such a rule should not exist in outer space. Such a hierarchy would recognize that the practical realities of outer space preclude the possibility of there being just one means of prescribing criminal jurisdiction. For the sake of legal and political certainty, meanwhile, a hierarchy would bring order and predictability to competing claims for jurisdiction.

Breaking with Convention: The Supreme Court's Approach to Separability in *Kabab-Ji v KFG*

Ryan Henry Ho*

ABSTRACT

This case note explores the recent ground-breaking decision rendered in Kabab-Ji, which saw the Supreme Court break with the rest of the world with regards to its interpretation of the separability of the arbitration agreement from the underlying contract (and the subsequent choice of law analysis that follows). It is ultimately argued that the English approach in Kabab-Ji is correct by tracing through the prior conceptions of the separability and demonstrating that the newer approach is more in line with what parties intend.

*LLM (Cantab) '23; LLB (LSE) '22. I would like to thank the LSE Law Review for their invaluable comments and feedback throughout. I am also grateful to my teammates, coaches, and opponents I met during the 29th Willem C. Vis Moot, without whom my ideas would have gone unrefined and untested. All errors remain my own.

INTRODUCTION

Long before the merits of a cross-border dispute can be examined, a common – and often lengthy¹ – dispute between the parties arises as to which court has jurisdiction and what law applies. Although parties attempt to pre-empt this by incorporating exclusive jurisdiction clauses, choice of law clauses, and arbitration agreements,² the question of choice of law can arise nevertheless. For example, when an arbitration clause is disputed, parties often challenge the law applicable to determine the validity thereof.

This problem is compounded by the now-trite principle that the arbitration agreement is presumed to be separable, such that its invalidity does not automatically follow from the alleged or actual demise of the underlying contract.³ How this ‘separability principle’ relates to the arbitration agreement, and how this affects the choice of law relating thereto, has in turn generated much controversy globally, with much of the debate concerning the strength of the separability presumption.⁴

While *Kabab-Ji* seemed like one among many such disputes regarding the law applicable to the arbitration agreement, the Supreme Court’s decision

¹ *Vedanta Resources PLC and another (Appellants) v Lungowe and others (Respondents)* [2019] UKSC 20 [7]; *VTB Capital plc v Nutritek International Corp* [2013] 2 AC 337 [82]-[83].

² Queen Mary University of London, ‘2021 International Arbitration Survey: Adapting arbitration to a changing world’ (*Queen Mary University of London*, 2021) <https://arbitration.qmul.ac.uk/media/arbitration/docs/LON0320037-QMUL-International-Arbitration-Survey-2021_19_WEB.pdf> accessed 19 November 2022, 5-6; Ardavan Arzandeh, ‘Exclusive Jurisdiction Clauses in International Trust Deeds’ (2021) 41 *Legal Studies* 527; Deyan Draguiev, ‘Unilateral Jurisdiction Clauses: The Case for Invalidity, Severability or Enforceability’ (2014) 31(5) *Journal of International Arbitration*, 19-45.

³ Gary Born, *International Commercial Arbitration* (3rd edn, Kluwer Law International 2021) 375; Margaret L. Moses, ‘The Arbitration Agreement,’ *The Principles and Practice of International Commercial Arbitration* (3rd edn, Cambridge University Press 2017) 21; Nigel Blackaby and others, *International Arbitration* (6th edn, Oxford University Press 2015) 104-107; Emmanuel Gaillard and John Savage (eds.), *Fouchard, Gaillard, Goldman on International Commercial Arbitration* (Kluwer Law International 1999) 197, 210.

⁴ Born (n 3) 383-385, 387-389.

overturned orthodox thought by disregarding the separability presumption altogether.

This essay will first examine the relevant facts and legal principles emerging from the Supreme Court's decision. Then, after exploring the differing interpretations of the separability principle, it will be concluded that the UK's approach is wholly compatible with the separability principle as it should be properly understood. Further, the Court's treatment of the choice of law clause and how it interacts with the separability principle of the arbitration agreement is more in line with what parties intend. This is for three reasons. First, the broad interpretation of separability goes further than necessary and is conceptually unsound. Second, the principle of separability does not actually require the premise that a governing law clause can never, without more, act as an express choice extending to the arbitration agreement as well. Finally, while this approach would be more beneficial and adhere to what parties intended, given that arbitral tribunals look at the entire corpus of law around the world, the *Kabab-Ji* approach requires more support globally before it can be relied on in tribunals outside England.

I. THE RELEVANT PRINCIPLES IN KABAB-JI V KFG

Kabab-Ji, a Lebanese company specialising in Lebanese and Middle Eastern cuisines, signed a Franchise Development Agreement ('FDA') with Al Homaizi, allowing the latter to operate a franchise in Kuwait.⁵ The relevant clauses of the FDA were as follows:

Art. 1: This Agreement consists of the foregoing paragraphs, the terms of agreement set forth herein below, the documents stated in it, and any effective Exhibit(s), Schedule(s) or Amendment(s) to the Agreement or to its attachments which shall be signed later on by both Parties. It shall be construed as a whole and each of the documents mentioned is to be regarded as an integral part of this Agreement and shall be interpreted as complementing the others.

⁵ *Kabab-Ji SAL (Lebanon) v Kout Food Group (Kuwait)* [2021] UKSC 48 [3].

Art. 14: any dispute ... with respect to any issue arising out of or relating to this Agreement or breach thereof [shall be settled in the International Chamber of Commerce, with the seat in France].

Art. 15: This Agreement shall be governed by and construed in accordance with the laws of England.⁶

Al Homaizi was then acquired by Kout Food Group ('KFG').⁷ In a subsequent dispute, Kabab-Ji brought KFG to arbitration under Art. 14 of the FDA.⁸ KFG attempted to contest that the tribunal did not have jurisdiction over themselves, as the arbitration agreement only relates to any disputes between Al Homaizi and Kabab-Ji, *not* KFG.⁹

The tribunal held that the law applicable to the arbitration agreement was French law and KFG was bound by the arbitration agreement, ultimately finding that KFG was in breach of the FDA.¹⁰ In parallel proceedings, while KFG sought to annul the award at the seat of arbitration in France, Kabab-Ji attempted to enforce it in England.¹¹

This is where the courts diverge. Prior to the judgment by the UK Supreme Court, the Cour d'Appel found no express choice of law applicable to the arbitration agreement. Thus, absent any choice of law, they found that French law, being the law of the seat of arbitration, governed the agreement.¹² Across the Channel, the Court of Appeal in England instead not only held that it was English law that applied, but it did so as a matter of express choice.¹³ Although Flaux LJ reaffirmed the presumption that an express choice of law clause cannot be an express choice extending to the arbitration agreement as well,¹⁴ this presumption was overturned on these facts by the terms of the FDA.

⁶ *ibid* [37].

⁷ *ibid* [4].

⁸ *ibid* [5].

⁹ *ibid* [20].

¹⁰ *ibid* [6].

¹¹ *ibid* [8].

¹² *ibid* [9]; Cour d'appel de Paris (2020) Case n°17-22943 [26], [47].

¹³ *Kabab-Ji S.A.L. (Lebanon) v Kout Food Group (Kuwait)* [2020] EWCA Civ 6 [62].

¹⁴ *ibid* [66].

The use of the phrase ‘This Agreement’ in Art. 1 made clear that it was intended to encapsulate all the terms of the contract; the same phrase, ‘This Agreement’, was also used when determining the governing law in Art. 15. When read together, an express choice of law for the underlying contract also expressly included the arbitration agreement.¹⁵

The Supreme Court however went a step further: they unanimously held that Art. 15 *alone* as ‘ordinarily and reasonably understood’ should include all terms of the contract *including* the arbitration agreement,¹⁶ making no mention of separability in its judgment. Subsequent to the Supreme Court’s decision, the Cour de Cassation affirmed the decision of the Cour d’Appel, leaving the two jurisdictions at an irreconcilable crossroads.¹⁷

II. SEPARABILITY

To appreciate why the Supreme Court’s decision was so surprising, two concepts must be understood: the rule on how to determine what law applies to an arbitration agreement, and the separability principle.

The rule to determine what law applies to an arbitration agreement is found in Article V.1(a) of the Convention on the Recognition and Enforcement of Foreign Arbitral Awards (‘New York Convention’). Although strictly dealing with the validity of arbitration awards, it has been interpreted by courts worldwide as also being the choice of law rule to determine what law applies to the arbitration clause.¹⁸ It lays out a three-stage test. First, whether the parties made an express choice of law applying to the arbitration agreement. Second, failing that, whether there were any indications thereon. Third, if neither were present, then the default is the law where the award is made, which is the law of the seat.

On the other hand, the separability principle, which forms the cornerstone of arbitration, states that the validity of the arbitration agreement is

¹⁵ *ibid* [62].

¹⁶ *Kabab-Ji UKSC* (n 5) [39].

¹⁷ Cour de Cassation (2022) n°20-20260.

¹⁸ *ibid* [26]; *Enka v Chubb* [2020] UKSC 38 [128]; *Kabab-Ji Cour d’Appel* (n 12); Gaillard and Savage (n 3) 218-219; Born (n 3) 529.

assessed separately from the underlying contract.¹⁹ Without it, parties could easily escape from their obligation to arbitrate by alleging that the underlying contract was void. As a result, courts have found that different laws can apply to the arbitration clause and the rest of the contract.

The way in which the separability principle affects the rule in Art.V(1)(a) has proven difficult. First, both common and civil jurisdictions agreed that a general choice of law clause cannot, without more, also be an express choice of law for the arbitration agreement.²⁰ However, whether it could constitute an implied choice of law was less clear. The case law is divided across the world,²¹ with two interpretations which will be referred to as the ‘narrow’ and the ‘broad’. While civil law jurisdictions endorsed the ‘broad’ interpretation, common law jurisdictions have typically supported the ‘narrow’ view.²²

The ‘narrow’ interpretation states that the arbitration agreement must be subject to a separate choice of law analysis from the underlying contract for the purpose of determining its validity.²³ The arbitration agreement can therefore, but not necessarily, be subject to a different law from that of the underlying contract.

In the UK, the Court of Appeal in *Sulamérica* held that the separability principle only presumes that parties intend for their dispute resolution mechanisms to remain effective when the underlying contract would be void.²⁴ Otherwise, if an arbitration tribunal finds that the underlying contract is invalid, the clause which conferred jurisdiction on that same tribunal is also ineffective. The tribunal would then paradoxically have no authority to declare its lack of jurisdiction. Conversely, if the fate of the arbitration agreement is automatically tied to the underlying contract, then any party could simply avoid an arbitration agreement by impugning the existence of the underlying contract. The ‘narrow’

¹⁹ Born (n 3) 376; Blackaby and others (n 3) 106; Moses (n 3) 21.

²⁰ Born (n 3) 375-376; *BNA v BNB* [2019] SGHC 142 [17]; *BCY v BCZ* [2016] SGHC 249; *Sulamérica CLA Nacional de Seguros S.A. v Enesa Engenharia S.A.* [2012] EWCA Civ 638 [26]-[27]; ‘Cyprus Case’ OGH 3 Ob 153/18y; ‘Transformer Case’ 18 OCg 6/18h; *Kabab-Ji Cour d’Appel* (n 12).

²¹ *Kabab-Ji UKSC* (n 5) [32]; *Enka v Chubb* (n 18) [55].

²² Born (n 3) 378-379.

²³ *Enka v Chubb* (n 18) [61].

²⁴ *Sulamérica v Enesa Engenharia* (n 20) [26].

conception therefore states that it should be presumed that the arbitration agreement is separable, which must give way to any contraindications.²⁵ However, even under the ‘narrow presumption’, there were doubts as to whether an express choice of law could serve as an *implied* choice of law for the arbitration agreement. While *Enka v Chubb* answered this in the affirmative, its applicability was limited only to the UK’s domestic choice of law rules,²⁶ and still preserved the notion that an ‘express choice of law’ regarding the arbitration clause can only exist if it is found within the clause itself.²⁷

Under the ‘broad’ interpretation, the separability principle instead requires that the arbitration agreement be autonomous for all purposes by operation of law.²⁸ Consequently, where a contract only has an express choice of law to the underlying contract, it cannot – without more – be an express choice and rarely be an implied choice of law for the arbitration agreement.²⁹ Instead, the law of the seat applies as either an implied choice or the default option under Art. V(1)(a) of the New York Convention.³⁰

In a way then, *Kabab-Ji* was both unsurprising and shocking at the same time. On one hand, this was merely a direct extension of the principles enunciated in *Enka* to Art. V(1)(a). On the other hand, *Kabab-Ji* did not treat the express choice of law of the main contract as an implied choice for the arbitration agreement, but an express choice as well.³¹ Further, unlike the domestic context of *Enka*, *Kabab-Ji* has to interact and contend with international jurisprudence on Art. V(1)(a). Nevertheless, it will be argued that the Supreme Court’s judgment was ultimately a better interpretation of the separability principle with reference to its purpose.

²⁵ *Enka v Chubb* (n 18) [60].

²⁶ *ibid* [27], [129]. In that case, it must also be noted that there was *no* express choice of law clause for the general contract.

²⁷ *ibid* [193].

²⁸ Judgment of 7 May 1963, *Ets Raymond Gosset v. Carapelli*, JCP Ed. G 1963, II, (French Cour de Cassation civ. 1e); Tribunal fédérale [TF] Sept. 2, 1993, DFT 119 II 380, 384; Born (n 3) 378.

²⁹ Gaillard and Savage (n 3) 212; *Kabab-Ji Cour de Cassation* (n 17).

³⁰ Born (n 3) 547-549; *Judgment of 28 September 1995*, XXII Y.B. Comm. Arb. 762, 765 (Rotterdam Arrondissementsrechtbank) (1997).

³¹ *Kabab-Ji UKSC* (n 5) [39].

III. DISPROVING THE BROAD INTERPRETATION OF SEPARABILITY

The broad interpretation of separability argues that the arbitration agreement is seen as ‘autonomous’ or independent for all purposes.³² This is an approach particularly endorsed by the French courts, where *Gosset* held that the arbitration agreement has ‘full legal autonomy’.³³ However, this interpretation must be rejected. Not only is this conceptually unsound, but its immutability also means that it can go against what the parties intended in fact. This is especially problematic as arbitration is founded on the consent of both parties.³⁴ Despite this, however, the broad approach has stood unquestioned for the past decades in civil law jurisdictions.³⁵

Historically, there have been two ways to justify the separability of the arbitration agreement: its ‘procedural’ or ‘judicial’ nature, as well as it reflecting the will of the parties. However, neither requires the total autonomy of the agreement.

With regards to the theory that the arbitration agreement is somehow an ancillary ‘procedural’ contract, this is not a settled matter and, in any case, does not require the total autonomy of the clause. While the classification of the arbitration agreement has been contested, what is indisputable is that its *formation* still concerns normal contractual principles. Rather, the so-called procedural elements of the contract concern, *inter alia*, the appointment of arbitrators and the admission of evidence. Similarly, and as mentioned below, the seat only has the power to apply its procedural rules on the fairness of the tribunal and its composition, *not* its substantive rules on contract formation. In both cases therefore, the purported *sui generis* nature of the arbitration agreement is insufficient to warrant its total autonomy in all cases.

On the other hand, the argument that the broad conception reflects the ‘will of the parties’ is also inaccurate. This justification has been deployed by

³² Born (n 3) 376-377; Gaillard and Savage (n 3) 216.

³³ *Ets Raymond Gosset v. Carapelli* (n 28).

³⁴ Born (n 3) 251; Blackaby and others (n 3) 71.

³⁵ Born (n 3) 376; Gaillard and Savage (n 3) 206; Y Derains 'Les Tendances de la Jurisprudence Arbitrale Internationale'. (1993) 120 Journal du Droit Intl 829; *Ets Raymond Gosset v. Carapelli* (n 28).

civil law jurisdictions such as France and Germany by reference to their own jurisprudence. The French courts, for example, declared that an agreement to arbitrate is valid by the will of the parties ‘without any reference to the law governing that agreement’, which they claim is a principle of international law. However, while this view has its merits and demerits, there is no such international law which supports this ‘delocalised’ theory of arbitration; instead, as Moses points out, the delocalised approach has not found much favour across the world.³⁶

Gaillard & Fouchard nevertheless attempt to justify the broad conception by stating that it reflects the will of the parties:

[as] arbitration rules derive their authority from the intentions of the parties ... where parties have referred to arbitration rules which enshrine the principle of the autonomy of the arbitration agreement, those parties are presumed to have intended that the arbitration agreement be treated separately from the main contract.³⁷

However, this assumes that the arbitration agreement is autonomous – itself a contentious claim – as well as the fact that the parties are aware of this. Reasonable businesspeople, and indeed as a matter of ordinary contractual interpretation, would intend for a single choice of law clause – with ‘this Contract’ or similar language – to encompass the entire contract, as they expect their contracts to be internally consistent. Since separability is a specific legal fiction pertaining to arbitration agreements which is not known to businesspeople, it is difficult to accept that the principle should apply to insulate the arbitration clause from their choice of law clause when they had given little or no thought to separability at all. Thus, the notion that the broad conception somehow reflects the ‘will of the parties’ becomes illusory. Nonetheless, the Cour de Cassation has, in their judgment, repeated the reasoning that ‘by virtue of a substantive rule of the law of international arbitration’, an arbitration

³⁶ Moses (n 3) p 65; Born (n 3) 1674-1675; Blackaby and others (n 3) 179-182; L. Collins (ed.), *Dicey, Morris and Collins on The Conflict of Laws* (16th edn, Sweet & Maxwell) 16-008; *Star Shipping AS v. China National Foreign Trade Transp. Corp.* [1993] 2 Lloyd’s Rep. 445, 452; Judgment of 20 December 1993, *Municipalité de Khoms El Mergeb v. Dalco*, 1994 Rev. Arb. 116 (French Cour de Cassation Civ. 1).

³⁷ Gaillard (n 18) 199-200.

agreement is legally independent of its underlying contract, assessed by reference to the will of the parties without the need for any state law.³⁸

The broad conception is made more unattractive when considering that it redirects focus to the law of the seat instead. This is done by two ways: either by claiming that the express choice of seat is important enough to displace or militate against a single choice of law clause, or by claiming that a difference in seat and location of choice of law is enough to justify resorting to the default option under Art. V(1)(a). However, this exaggerates the importance of the seat. Art. V(1)(a) states that the default law of the seat is only triggered when ‘failing any indication thereon’. This is a low bar, which a clear and uncontested choice of law clause should satisfy by virtue of it being evidence that parties have given thought to the matter. Conversely, the argument that the arbitration agreement only has power by the seat’s permission is also increasingly outdated, especially given how many signatories there are to the New York Convention. Rather, the better view is that parties freely agree to arbitration, and that states respect party autonomy by giving effect to what the parties agreed. Additionally, the argument that the seat is where the obligation to arbitrate is ‘performed’ does not hold weight. There is no obligation that the seat of arbitration be where the arbitration is physically heard, as evidenced by the fact that many arbitrations now allow online hearings due to the pandemic. Nowadays, the seat is merely the legal place where the award originates, which in turn entitles the seat to only apply its procedural laws in discharging their supervisory jurisdiction over the arbitration. The function of the law of the seat as a last-resort option also demonstrates this point: it is only when there is a total absence of party choice that courts and tribunals resort thereto.

Thus, the broad separability principle is wholly unnecessary for what it purports to do, as well as overemphasises the importance of the seat. Per Gaillard & Fouchard, the consequence of the autonomy of the arbitration agreement is that its validity is not affected by that of the contract, and that it can be governed by a different law from the main contract.³⁹ However, and as will be seen below, both are achieved via the narrow separability principle without the need for such an expansive reading.

³⁸ *Kabab-Ji Cour de Cassation* (n 17) [7].

³⁹ Gaillard and Savage (n 3) 209.

IV. REMOVING UNNECESSARY PRESUMPTIONS

While the narrow conception is better in that it does not force the independent and autonomous nature of the arbitration agreement, it nevertheless presumes that an express choice of law cannot automatically apply as an express choice for the arbitration agreement. However, as seen in *Kabab-Ji*, this presumption is unnecessary. This is because the *kompetenz-kompetenz* principle – that the tribunal has power to rule on its jurisdiction – has become so prevalent that it has force on its own.

First, the objective of the narrow interpretation is to protect the validity of the arbitration agreement from the demise of the underlying contract. However, this does not necessarily relate to applicable law. To illustrate this, the initial tribunal which was called upon to settle the dispute in *Kabab-Ji* was the International Chamber of Commerce, whose rules state:

Article 6(9): Unless otherwise agreed, the arbitral tribunal shall not cease to have jurisdiction by reason of any allegation that the contract is non-existent or null and void, provided that the arbitral tribunal upholds the validity of the arbitration agreement. The arbitral tribunal shall continue to have jurisdiction to determine the parties' respective rights and to decide their claims and pleas even though the contract itself may be non-existent or null and void.

Under these rules, if KFG instead sought to contest that the entire underlying contract was void, regardless of whatever law they claim it was invalid under, this would still be irrelevant. This is because the separability principle is now so widely accepted that it essentially operates *prior to* choice of law.⁴⁰ In other words, rather than recognising that separability derives force from a specific national court, a better view is that it has become such a necessary part of arbitration that it has force on its own. Indeed, it is impossible for a tribunal to not provide for some version of separability, because it would effectively have no jurisdiction, as seen by the fact that rules providing for

⁴⁰ Gaillard and Savage (n 3) 199.

kompetenz-kompetenz can be found in those of other major tribunals.⁴¹ Therefore, the presumption that the validity of the arbitration agreement and underlying contract can still be upheld without the tribunal having to determine the law applicable to the arbitration agreement.

By doing away with the presumption of separability and treating express choice of law as including the arbitration agreement by default, this approach would firstly be more faithful to the text of the contract and what parties intended. While what the parties intended during the contractual negotiation phase may be subjective and differ, the contract is what conclusively reflects the *shared* intentions of the parties. Given that it is more accurate to work from the premise that businesspeople do not intend for a single line in their dispute resolution clauses to escape their broader choice of law,⁴² the requirement should be for the disputing party to positively prove that the parties intended that a separate law should apply to the arbitration agreement. This is also helpful where a contract is concluded orally. In such contexts, while parties may have agreed on the applicable choice of law, they may neglect to specify or even consider the choice of law applicable to the arbitration clause. In such scenarios, the better explanation is that the parties intend for their choice of law clause to govern the entire arbitration agreement. This is also more persuasive, as the alternative is applying the law of the seat, which may have never been in contemplation. In any case, even where parties expressly choose their seat, they do not intend for the law of the seat to determine the substantive validity of the arbitration clause; the factors which influence the choice of seat are an established reputation, greater support from the judiciary, and neutrality of the legal system.⁴³ Clearly, these considerations do not concern the substantive law of the seat, but rather its procedural laws such as those on evidence, replacing arbitrators or challenging awards. Conversely, it is difficult to see how choosing a seat can be seen as accepting its substantive law.⁴⁴

One concern is that the abolishment of the presumption would make it difficult to resist the conclusion that a choice of law clause could be anything but an express choice applying to the arbitration agreement. However, such

⁴¹ LCIA Arbitration Rules (1 October 2020) art 23.2; HKIAC Administered Arbitration Rules (1 November 2018) art 19.2.

⁴² *Enka v Chubb* (n 18) [39], [43].

⁴³ Queen Mary University of London (n 2) 7, 8.

⁴⁴ Blackaby and others (n 3) 166.

concerns are unfounded. Suppose the entire contract including the arbitration and choice of law clauses were negotiated as inseparable and then contained within one document; the conclusion that the choice of law also applies to the arbitration agreement would be fair, as the parties clearly contemplated both the arbitration clause and choice of law clause together. Conversely, there are also scenarios when extending an express choice of law to the arbitration clause would be less tenable, such as where the parties clearly did not contemplate arbitration at the time of concluding the contract but opted to include it subsequently.

Another counterargument is that one should not adopt the viewpoint of reasonable businesspeople when interpreting arbitration clauses, but rather the view of lawyers who draft these clauses. Indeed, if any lawyer were to open a textbook or practitioners' guide (pre-*Kabab-Ji*), they could assume that their express choice of law for the underlying contract would not include the arbitration agreement; they may even assume that it is instead the law of the seat that applies.

However, this takes too artificial a view of the parties. First, if all lawyers knew of separability, they would have also been aware of the ongoing debate on whether the choice of law clause can be an implied choice as well. They would have then properly written out two choice of law clauses for the arbitration agreement and the underlying contract separately. Regardless, lawyers are agents of the parties, and their interpretation, mistakes, omissions, or deviations from the party's intentions should not easily factor into interpretation. This is also in line with English principles of contractual interpretation: contractual language is to be interpreted by what a reasonable person with all the background knowledge would understand the words to mean, with a view to the natural and ordinary meaning of the words, disregarding subjective evidence of parties' intentions.⁴⁵

Finally, this change would not radically alter outcomes. In both civil and common law jurisdictions, it is increasingly held that the law of the underlying contract, rather than the law of the seat, is an implied choice for the

⁴⁵ *Arnold v Britton* [2015] UKSC 36 [15]-[23].

arbitration agreement.⁴⁶ Nevertheless, recognising the choice of law as including the arbitration agreement would not only save time and litigation costs, it would also mean that the conceptual lens by which the priority of separability is interpreted going forward is corrected.

In summary, the separability presumption does not even require the presumption that an express choice of law is barred from being an express choice applying to the arbitration agreement as well. Instead, treating an express choice of law as *including* the arbitration agreement *ab initio* – barring any obvious contraindications – would be more faithful to what the parties intended. Thus, the decision in *Kabab-Ji*, which followed the narrow interpretation while disregarding the unnecessary presumption, was ultimately satisfactory.

V. PRACTICAL ISSUES & CONCLUSION

Some practical issues remain, however. As tribunals are expressly not bound by one system of law,⁴⁷ *Kabab-Ji* is not a decisive authority, especially when it stands alone in the face of overwhelming and opposing authority from other jurisdictions.⁴⁸ Parties therefore face an uphill battle convincing tribunals to adopt the view in *Kabab-Ji*, especially if the arbitrators are from civil jurisdictions.

For now, parties should always have two choices of law clauses, indicating separately the applicable law to the contract and the arbitration clause even if they would be the same. The risks of not doing so can mean that an award can either be set aside at the seat or not recognised at the place of enforcement. The facts of the current dispute are illustrative: while the award has been upheld at the seat, the Supreme Court refused the recognition of that award in the jurisdiction where the losing party had their assets, rendering it moot. Unsurprisingly, the Cour de Cassation affirmed the decision of their lower court, meaning other courts may be less hesitant to refuse enforcement of the award in their own jurisdictions as the award has been affirmed at the seat.

⁴⁶ *BNA v BNB and Another* [2019] SGCA 84; *BCY v BCZ* (n 20); ‘Cyprus Case’ (n 20); ‘Transformer Case’ (n 20).

⁴⁷ Born (n 3) 313.

⁴⁸ *Ets Raymond Gosset v. Carapelli* (n 28); ‘Cyprus Case’ (n 20); ‘Transformer Case’ (n 20); *Kabab-Ji Cour d’Appel* (n 12).

In conclusion, the Supreme Court's decision in *Kabab-Ji* is both more faithful to the language of the contract and better accords with what businesspeople intend. By overturning the heretofore unquestioned assumption that an express choice of law for the contract can never be an express choice to the arbitration agreement, the Supreme Court's approach now accurately reflects the purpose of the separability principle without going unnecessarily further. As the judgment did not directly address separability, this essay attempts to elucidate the conceptual foundations and proper scope of the principle following the *Kabab-Ji* decision. Ultimately, it is hoped that other jurisdictions will take inspiration from *Kabab-Ji* and arrive at the same conclusion that the Supreme Court did: that a legal fiction should not take precedence over what the parties intended in fact.

Neuro-Doping, tDCs and Chess — Are WADA's Regulations under threat?

Ryan Nathan Au*

ABSTRACT

The recent emergence of cognitive enhancement techniques, including non-invasive brain stimulation (NIBS), has reignited discussions on the serious concerns regarding the World Anti-Doping Agency's (WADA) outdated anti-doping regulations. Hitherto, much has been written from the perspective of drugs, physical enhancements and physical sports, but little has been analysed through the lens of NIBS/cognitive enhancement/mind sports. Using the example of transcranial direct current stimulation (tDCS), a form of NIBS, in professional chess as a case study, this article contends that WADA's anti-doping regulations are, paradoxically, both under- and over-exclusive in part as a result of physical sports scandals in the 1980s. This article comprises three Sections. The first Section identifies four key problematic areas of WADA's anti-doping regulations, including the failures of WADA's 'two-thirds rule' as well as the failure to distinguish between physical and mind sports, before engaging with the application of tDCS in chess in order to illustrate the existence of deep-rooted defects within WADA's framework. Section II argues that WADA's deontological approach is unjust in the context of cognitive enhancement, before proceeding to deconstruct and question the validity of the key underlying ethical principles that shape WADA's policy. Finally, Section III offers a vision of a decentralised regime which gives priority to the health of the athlete, pays regard to actual enhancements and appreciates the nuances of individual sports. Such a regime would create a more robust and equitable regulatory system.

* LLB (LSE) '22. The author is most grateful to Professor Emily Jackson and Dr Cressida Auckland for their supervision and guidance. The author would also like to thank his parents for their support as well as the editorial team.

ABBREVIATIONS

CAS	Court of Arbitration for Sport
EEG	Electroencephalography
EPO	Electroencephalogram
FIDE	World Chess Federation (<i>Fédération Internationale des Échecs</i>)
FINA	World Swimming Federation (<i>Fédération Internationale de Natation</i>)
FIVB	International Volleyball Federation (<i>Fédération Internationale de Volleyball</i>)
IPC	International Paralympic Committee
IOC	International Olympic Committee
LTM	Long-term memory
NCAA	National Collegiate Athletic Association
MDD	Major depression disorder
NADOs	National anti-doping organisations
NIBS	Non-invasive brain stimulation
PLIS	Prohibited List International Standard (effective as of 1 January 2022)
tACS	Transcranial alternating current stimulation
tDCS	Transcranial direct current stimulation
tRNS	Transcranial random noise stimulation

tMS	Transcranial magnetic stimulation
WADA	World Anti-Doping Agency
WADC	WADA's Anti-Doping Code 2021
World Athletics	International Amateur Athletics Federation (formerly known as International Amateur Athletic Federation)

INTRODUCTION

The etymological origins of ‘sport’¹ come from the old English word ‘disport’, meaning a ‘diversion from serious duties; relaxation, recreation, entertainment, amusement’.² Over the centuries, however, the concept of sport has evolved into a much more serious endeavour. The Olympic motto ‘*Citius, Altius, Fortius — Communitur*’ or ‘faster, higher, stronger — together’ embodies the notion of competition, the use of comparatives indicative of a notion of superiority over one’s competitors, that is, a competition *for victory*. At the same time, athletes must play by the rules: they must not achieve victory by any means possible and they must not, in any way, seek to unfairly gain advantage over other athletes. In other words, they must not *cheat*. It is against this background that ‘doping’ — the act of taking illicit substances/methods for the purpose of performance enhancement — is subject to intense regulation by WADA, a foundation set up in 1999 under the initiative of IOC with more than 650 signatories to WADC, its anti-doping Code, which include international sports federations, NADOs and IPC³. What emerges is a tension between the pressures of victory — exacerbated by contemporary realities of funding pressures, national pride and sponsorships — and a duty to uphold the fundamental rationale of WADC — the ‘spirit of sport’.

Scope and Structure of this Article

This article seeks to explore the multiple fundamental deficiencies that plague WADA’s approach to doping. These deficiencies include, in particular, issues associated with an outdated concept of the ‘spirit of sport’ for anti-doping regulations and a *duty*-based regulatory framework that punishes the *intention* to gain an unfair advantage, as opposed to the *outcome* of actually gaining an unfair advantage. Concerningly, such an intention can be further stretched to capture circumstances where the required state of mind does not even exist, as

¹ ‘Sport, *n.*1: Oxford English Dictionary’ (OED.com, 2022)
<<https://www.oed.com/viewdictionaryentry/Entry/187476>> accessed 26 April 2022.

² *ibid.*

³ WADA, ‘Code Signatories’ (World Anti-Doping Agency November 8, 2022)
<<https://www.wada-ama.org/en/what-we-do/world-anti-doping-code/code-signatories>> accessed November 26, 2022.

this article will illustrate in Section 1.1. These deficiencies are exacerbated by the recent emergence of NIBS techniques such as tDCS, which pose new issues, especially in the context of cognitive enhancement. The emergence of NIBS techniques has sparked debate over their future regulation and (re)ignited past discussions concerning the adequacy of WADA's criteria which currently regulate *some* forms of cognitive enhancement, but not all. Debates on NIBS add a new dimension to the existing chorus of dissatisfaction with WADA's regime by bringing to the fore the problem of its 'all-inclusive' approach. Such an approach embraces indiscriminate application of doping regulations to *all* sports, regardless of the nuances of each individual sport, backed by the omnipotent spirit of sport as a philosophy and a regulatory criterion. The time is indeed ripe for WADA to seriously reflect on its increasingly untenable position both with regards to upholding its current anti-doping regulations as well as to its own existence as a governing body. Additionally, since the issue of doping has far-reaching implications for professional athletes and other stakeholders, a more considered approach by WADA could help orient economic and political discussions that will ultimately help shape important policy decisions.

This article will be divided into three sections. Section I, Part A provides context by identifying four key problematic areas of WADA's anti-doping regulations contained in WADC. WADC is a lengthy document with a broad coverage ranging from testing to hearings. This article specifically focuses on the compilation of PLIS pursuant to Article 4 of WADC. Section I, Part B will analyse whether new NIBS technologies can, using tDCS and chess as examples, be banned under the current WADA framework. Section II will then deconstruct WADA's anti-doping regulations in light of those findings. Finally, Section III will conclude by offering alternative pathways going forward, which include placing the athlete's health at the centre of future regulations.

Justifying the Scope

Before examining WADA's regulations, it is pertinent to first explain why tDCS and chess are used as analytical tools.

a) Definitions

tDCS is a form of non-invasive neurostimulation that involves ‘the placement of electrodes on the human scalp for the purpose of stimulating specific brain functions’.⁴ Low current is briefly transmitted through the skull to alter ‘cortical excitability and activity dependent on the current flow direction through the target neurons’.⁵ There is a growing body of literature that suggests that tDCS is ‘safe and effective when used for therapeutic purposes’.⁶

Before examining further *why* tDCS was selected, it is pertinent to clarify the definitions of key terms, specifically, ‘cognition’, ‘cognitive enhancement’, ‘NIBS’ and ‘neuro-doping’. Cognition relates to the organisation, acquisition, selection, representation and retention of information, and cognitive enhancement may be directed at improving any one of these core faculties.⁷ According to Sandberg, a huge range of such enhancement methods exists⁸ from relatively non-controversial educational techniques aimed at improving efficiency (such as mnemonics, encoding strategy and speed reading) and approaches aimed at improving health generally (through sleep, immune, exercise or improved nutrition)⁹ to more controversial techniques, such as drug stimulants (e.g., caffeine and nicotine) and genetic modifications.

Neuro-doping, on the other hand, refers to ‘direct interventions into the brain’¹⁰ and typically takes the form of electrical stimuli being fired into specific areas of the brain — ‘as opposed to other parts of the body’¹¹ — to stimulate a reaction, whether it is increased reaction time or memory retention and recall. Such technology has hitherto been used in a clinical setting, such as

⁴ Colton D Smith, ‘The Electric Brain: Do-It-Yourself Healthcare With Transcranial Direct Current Stimulation’ (2017) 5 *Journal of Responsible Innovation* 88.

⁵ AJ Woods and others, ‘A technical guide to tDCS, and related non-invasive brain stimulation tools’ (2016) 127 *Clinical Neurophysiology* 1031.

⁶ Smith (n 4) 89.

⁷ Nick Bostrom and Anders Sandberg, ‘Cognitive Enhancement: Methods, Ethics, Regulatory Challenges’ (2009) 15 *Science and Engineering Ethics* 312.

⁸ *ibid.*

⁹ Ulric Neisser, ‘Rising Scores on Intelligence Tests’ (1997) 85 *American Scientist* 440; Phillip D Tomporowski, ‘Effects of acute bouts of exercise on cognition’ (2003) 112 *Acta Psychologica* 297.

¹⁰ Alexandre Erler, ‘Neuro-Doping and the Value of Effort in Endurance Sports’ (2020) 14(2) *Neuroethics* 1.

¹¹ *ibid.*

through tMS, to treat, for example, MDD.¹² Two essential points must be highlighted at the outset. First, neuro-doping differs from cognitive enhancement in one key aspect: the term ‘doping’ incorporated in ‘neuro-doping’ implies that the use of such an enhancement is necessarily an ‘unauthorised means to increase performance in sport’.¹³ The current emerging cognitive-enhancing technologies, such as tDCS, tRNS, tACS and EEG, do not fall within the ambit of WADC (as yet) and it would accordingly be a misnomer to refer to these methods as ‘neuro-doping’. Indeed, it is the intention of this article to establish whether such methods should be regarded as ‘neuro-doping’ and thus regulated by WADA.

Secondly, there is ‘no sharp line between n[euro]-d[oping] and non-n[euro]-d[oping]’¹⁴ since traditional doping (e.g., through the use of anabolic steroids) can affect mental capacities and the use of neuro-doping can also affect physical ‘endurance and muscle strength’.¹⁵ Given the overlap, it is only a matter of degree to what extent neuro-doping can enhance cognition and/or physical capacities for a given sport, though methods directed at cognitive improvement are more likely to result in cognitive rather than physical enhancement. This point is particularly relevant to one of the assertions on the physical/mental divide discussed in Section 2.1. The correlation between cognitive enhancement, NIBS/neuro-doping and physical enhancement can be best depicted in the Venn diagram as follows.

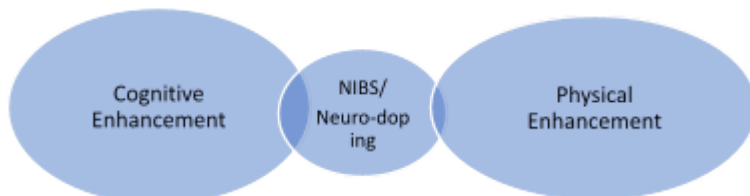


Figure 1: Venn diagram showing the correlation between neuro-doping and cognitive and physical enhancement, respectively.

¹² SL Garnaat and others, ‘Updates on Transcranial Magnetic Stimulation Therapy for Major Depressive Disorder’ (2018) 41 *Psychiatric Clinics of North America* 419.

¹³ Nick J Davis, ‘Neurodoping: brain stimulation as a performance-enhancing measure’ (2013) 43 *Sports Medicine* 649.

¹⁴ TS Petersen and K Lippert-Rasmussen, ‘Neuro-Doping and Fairness’ (2020) 14 *Neuroethics* 1179.

¹⁵ *ibid.*

b) Why tDCS

The reasons for the choice of tDCS are threefold. First, tDCS, as an emerging technology, has already shown its potential to impact both physical and cognitive performance. In terms of enhancement in physical sports, studies have shown that tDCS can help with accuracy in pistol shooting,¹⁶ basketball¹⁷ and golf putting,¹⁸ as well as enhance muscular force generation for sports requiring vertical jumping ability.¹⁹ Regarding cognitive enhancement, studies have suggested that tDCS can help with memory,²⁰ numerical magnitude processing,²¹ complex verbal problem-solving²² and three-dimensional mental rotation performance.²³ Davis argues that neuro-doping will become ‘a key technology for the future of sport and sports medicine’,²⁴ whilst Zhu Z et al. are of the view that ‘tDCS has strong potential to enhance athletic performance’.²⁵

Secondly, tDCS as a technique is already being distributed and used commercially, with companies like Halo Neuroscience showcasing professional

¹⁶ AM Kamali and others, ‘Transcranial Direct Current Stimulation to Assist Experienced Pistol Shooters in Gaining Even-Better Performance Scores’ (2019) 18 *Cerebellum* 119.

¹⁷ MA Friehs and others, ‘Electrify Your Game! Anodal tDCS Increases the Resistance to Head Fakes in Basketball’ (2019) 4 *Journal of Cognitive Enhancement* 62.

¹⁸ FF Zhu and others, ‘Cathodal Transcranial Direct Current Stimulation over Left Dorsolateral Prefrontal Cortex Area Promotes Implicit Motor Learning in a Golf Putting Task’ (2015) 8 *Brain Stimulation* 784.

¹⁹ E Lattari and others, ‘Can Transcranial Direct Current Stimulation Improve Muscle Power in Individuals with Advanced Weight-Training Experience?’ (2020) 34 *Journal of Strength and Conditioning Research* 97.

²⁰ NR Polizzotto, N Ramakrishnan, and RY Cho, ‘Is It Possible to Improve Working Memory With Prefrontal Tdcs? Bridging Currents to Working Memory Models’ (2020) 11 *Frontiers in Psychology* 1.

²¹ TU Hauser and others, ‘Enhancing Performance in Numerical Magnitude Processing and Mental Arithmetic Using Transcranial Direct Current Stimulation (tDCS)’ (2013) 7 *Frontiers in Human Neuroscience* 1.

²² C Cerruti and G Schlaug, ‘Anodal Transcranial Direct Current Stimulation of the Prefrontal Cortex Enhances Complex Verbal Associative Thought’ (2009) 21 *Journal of Cognitive Neuroscience* 1980.

²³ R Zhu, Z Wang, and X You, ‘Anodal Transcranial Direct Current Stimulation Over the Posterior Parietal Cortex Enhances Three-Dimensional Mental Rotation Ability’ (2021) 170 *Neuroscience Research* 208.

²⁴ Davis (n 13) 649.

²⁵ Z Zhu and others, ‘Commentary: “Brain-Doping,” Is It a Real Threat?’ (2019) 10 *Frontiers in Physiology* Article 1489.

tDCS devices and with evidence of tDCS use by Olympic athletes²⁶ and professional sports teams.²⁷ Along with increased accessibility and availability comes the concern of reporting bias. For example, Halo Neuroscience has been marketing the results of its own commissioned studies regarding tDCS.²⁸ Even though these studies fail to offer a holistic or accurate picture of the technology, it is entirely conceivable that athletes seeking a competitive advantage will nonetheless be lured by such aggressive advertising strategies and by what tDCS purportedly offers.²⁹ Given these campaigns, athletes may come to associate tDCS devices psychologically with safety, simplicity and affordability. It is therefore incumbent on WADA, as the coordinator of anti-doping efforts, to address without delay the many issues plaguing WADC which hinder proper and effective development of anti-doping regulations amidst the mounting concerns of tDCS usage in the sports world.

Thirdly, debates on tDCS can offer significant contributions to pre-existing literature on the acceptability of and issues with doping. Although ethical debates on cognitive enhancement in general are abundant, these have so

²⁶ 'US Athletes Using Brain Stimulation to Improve Olympic Performance' (RT International, 2022) <<https://www.rt.com/sport/354921-us-athletes-using-brain-stimulation/>> accessed 26 April 2022.

²⁷ Alex Hutchinson, 'For the Golden State Warriors, Brain Zapping Could Provide an Edge' (*The New Yorker*, 25 June 2016) <<http://www.newyorker.com/tech/elements/for-the-golden-state-warriors-brain-zapping-could-provide-an-edge>> accessed 20 April 2022.

²⁸ See, for example: Halo Neuroscience, 'Bihemispheric Transcranial Direct Current Stimulation with Halo Neurostimulation System Over Primary Motor Cortex Enhances Fine Motor Skills Learning in a Complex Hand Configuration Task' (2016); Halo Neuroscience, 'Bihemispheric Transcranial Direct Current Stimulation with Halo Neurostimulation System Over Primary Motor Cortex Enhances Rate of Force Development in an Isometric Lateral Pinch Force Task' (2016); Halo Neuroscience, 'Safety of Non-Invasive Brain Stimulation Delivered via the Halo Neurostimulation System in Healthy Human Subject' (2016).

²⁹ On inaccuracy, see, for example, on decrease in working memory and tDCS headsets: L. Steenbergen and others, "'Unfocus" on *foc.us*: commercial tDCS headset impairs working memory' (2015) 234 *Experimental Brain Research* 637. On lures of interventions, see B. Goldman and R. Klatz, *Death in the Locker Room* (Elite Sports Medicine Publications 1992).

far been mainly focused on its use in academia³⁰ or by military personnel,³¹ leading to Petersen's claim that the 'ethical analysis on the use or possible use of neuro-doping in sport is a neglected subject'.³² This vacuum is undesirable and, from a public policy perspective, how WADA chooses to regulate (or not to regulate) may send (unintentional) messages to aspiring athletes as to the nature of doping and the extent to which enhancement is acceptable. This article is motivated by the conviction that if debate on the regulation of tDCS in sports is delayed, regulation risks becoming merely reactionary to scandals. As Ritchie and Jackson have pointed out, 'instead of rational, proactive, and ethically based decision-making, "suits and administrators" reacting to politically charged events and shocks have pervaded the history of anti-doping policies'.³³ For example, the national consternation and embarrassment of Ben Johnson's stripped Olympic gold medal³⁴ directly led to the creation of the Canadian Centre for Ethics in Sport and their spirit of sport campaign.³⁵ This subsequently morphed into the very basis of WADA's entire regulatory framework. The issue with such regulation, however, is that it is influenced by the shock and emotional impact of scandals and not, as Ritchie and Jackson suggest, a balanced and 'ethical' decision-making process.³⁶ Therefore, the earlier the discourse, the better the regulation, and the more thorough and considered the resultant regime.

c) Why chess

The reasons for selecting chess as a case study for the effects of tDCS are fourfold. First, significant headway has been made in the cooperation between the governing body of chess FIDE and WADA. Being a signatory to WADC, FIDE has adopted WADA's anti-doping policy in full and, in 2015,

³⁰ Scott Vrecko, 'Just How Cognitive Is "Cognitive Enhancement"? On the Significance of Emotions in University Students' Experiences with Study Drugs' (2013) 4 *AJOB Neuroscience* 4-12.

³¹ TT Brunyé and others, 'A Review of US Army Research Contributing to Cognitive Enhancement in Military Contexts' (2020) 4 *Journal of Cognitive Enhancement* 453.

³² Thomas Søbirk Petersen, 'Sport, Neuro-Doping and Ethics' (2021) 14 *Neuroethics* 1.

³³ I Ritchie and G Jackson, 'Politics and "shock": reactionary anti-doping policy objectives in Canadian and international sport' (2014), 6 *International Journal of Sport Policy and Politics* 195.

³⁴ 'Johnson Stripped of Gold' (*BBC Sport*, 27 September 2013)

<<https://www.bbc.com/sport/av/athletics/24276046>> accessed 26 April 2022.

³⁵ Ritchie and Jackson (n 33)

³⁶ *ibid.*

published a comprehensive anti-doping code based on WADC and WADA's Models of Best Practice.³⁷ Amongst the various duties, FIDE must, under Article 18 of the FIDE anti-doping rules, periodically report to WADA on the compliance of FIDE's anti-doping policy with WADC. As such, any changes to WADA's policy will have subsequent knock-on effects on the regulation of doping in chess.

Secondly, chess is a uniquely suited sport for discussing cognitive enhancements since performance is comparatively easy to quantitatively assess. The universal system of performance rating in chess — the FIDE rating — allows the strength of players to be measured by reference to a predetermined algorithm that reflects performance in tournaments. Although this standardised unit of measurement does not directly illustrate *actual* cognitive performance, it is nonetheless a valuable tool that allows comparisons of skill levels to be drawn between players and with respect to the player themselves. Furthermore, since chess is a game with a finite number of moves, the quality of each move can be objectively assessed against how external factors, such as fatigue, might impact performance. Finally, with chess being a game with mathematical foundations comprising of patterns, aspects like opening theories and endgame strategies can be linked to certain areas of cognitive abilities, which lend themselves well to controlled research, such as detection of memory recall and pattern recognition. These features, especially measurability, are valuable attributes not necessarily present in all mind sports, such as Esports (which will be discussed later).

Thirdly, and relatedly, a study on the effects of psychopharmacological drugs on chess performance conducted in 2017³⁸ concluded that Ritalin and Modafinil can enhance chess performance by some 13–15%.³⁹ Although no equivalent research directly on the effects of tDCS on chess has been conducted, this study suggests that chess performance could be capable of

³⁷ FIDE Handbook

<<https://handbook.fide.com/files/handbook/FIDEAntiDopingRules.pdf>> accessed 26 April 2022.

³⁸ AG Franke and others, 'Methylphenidate, modafinil, and caffeine for cognitive enhancement in chess: A double-blind, randomised controlled trial' (2017) 27 *European Neuropsychopharmacology* 248.

³⁹ Dylan Loeb McClain, 'Proven: performance-enhancing drugs for chess' (*ChessBase*, 25 January 2017)

<<https://en.chessbase.com/post/proven-performance-enhancing-drugs-for-chess>> accessed 28 September 2022.

enhancement via external aid. In turn, it is possible to posit that similar conclusions might be drawn if tDCS is used. Studies have already shown that tDCS has potential to enhance qualities including information processing,⁴⁰ attention,⁴¹ alertness⁴² and memory⁴³ — four cognitive aspects which are particularly relevant to chess performance. Chess is therefore uniquely placed in any study on the potential applicability of tDCS.

Finally, the nature of chess as a mind sport raises unique questions as to our understanding of the spirit of sport. This concept, which has already been criticised for being excessively vague (which the analysis below will substantiate),⁴⁴ faces even more problems when applied in the context of chess. The spirit of sport criterion was created largely as a reactionary response to the controversies in *physical* sports, which means that it does not capture the differing values underpinning mind sports. Indeed, the values inherent to mind sports such as chess differ greatly from those of, for example, sprinting — an admittedly fine distinction to be explored in more detail in Section 1.3. Moreover, the spirit of sport criterion fails to appreciate the nuanced distinctions present between sports, even in the same category; for example, while reaction times are crucial to performance in Esports, good chess play is often equated with slow and measured calculations to avoid blunders.

With FIDE's commitment to anti-doping, the standardised measurement of chess performance, empirical data on tDCS studies, as well as the questions it raises about the spirit of sport, mean that chess lends itself very well as an appropriate sport through which to consider the regulation of neuro-interventions.

⁴⁰ R Abedanzadeh, S Alboghebish, and P Barati, 'The effect of transcranial direct current stimulation of dorsolateral prefrontal cortex on performing a sequential dual task: a randomized experimental study' (2021) 34 *Psicologia: Reflexão e Crítica*.

⁴¹ C Cosmo and others, 'A Randomized, Double-Blind, Sham-Controlled Trial of Transcranial Direct Current Stimulation in Attention-Deficit/Hyperactivity Disorder' (2015) 10 *PLoS One* 1.

⁴² LK McIntire and others, 'Transcranial direct current stimulation versus caffeine as a fatigue countermeasure' (2017) 10 *Brain Stimulation* 1070.

⁴³ L Huo and others, 'Transcranial Direct Current Stimulation Enhances Episodic Memory in Healthy Older Adults by Modulating Retrieval-Specific Activation' (2020) 2020 *Neural Plasticity* 1.

⁴⁴ See, for example, I Waddington and others, 'Recreational drug use and sport: Time for a WADA rethink?' (2013) 2 *Performance Enhancement and Health* 41.

I. SECTION I: WADA's REGULATIONS

Part A: Four Key Concerns

This Section will discuss four key themes with respect to WADA's current anti-doping regulations: two of these relate to the criteria for prohibiting a substance/method, and the other two relate to WADA's failure to observe various crucial distinctions. These four key concerns are by no means compartmentalised: issues relevant to one may be relevant to another and may even conflate.

Doping is defined in WADC as 'circumstances and conduct'⁴⁵ that violate the code (Article 1 of WADC). These range from the 'presence of *prohibited substances* in athletes' samples' (Article 2.1) to 'attempts on evading, refusing or failing to submit doping samples' (Article 2.3). Debate has hitherto been focused on the former and, as per Article 4.3, the inclusion of a substance on the prohibited list (PLIS) requires fulfilment of two of the three criteria (i.e., 4.3.1.1-4.3.1.3 of WADC), hereafter referred to as the 'two-thirds criteria'. This article is focused on criticising the two-thirds criteria set out in Article 4.3, which states that WADA shall consider the following when deciding if a substance or method ought to be included in PLIS:

4.3.1.1 Medical or other scientific evidence, pharmacological effect or experience that the substance or method, alone or in combination with other substances or methods, has the *potential* to enhance or enhances sport performance;
(Enhancement Criterion)

4.3.1.2 Medical or other scientific evidence, pharmacological effect or experience that the use of the substance or method represents an actual or *potential* health risk to the athlete;
(Health Criterion)

⁴⁵ WADC

<https://www.wada-ama.org/sites/default/files/resources/files/2021_wada_code.pdf > accessed 29 April 2022.

4.3.1.3 WADA's determination that the use of the substance or method violates the spirit of sport described in the introduction to the Code. (*Spirit of Sport Criterion*)

Of the three, the inclusion of the health criterion has been the least controversial. By any measure, the health of athletes should be at the core of any regulatory stance, and should as far as possible be protected — a point more fully explored in the final Section. What is more contentious, however, is the enhancement and the spirit of sport criteria. These will be discussed below, though it is worth highlighting here that the spirit of sport is *more than* just a criterion. The spirit of sport has been expressed as a fundamental rationale for anti-doping ('spirit of sport clause') enshrined in the preamble of WADC. This term and the many perspectives it purportedly represents are expanded upon in Section 1.2. Of relevance and worth stating for now is that lively debates have emerged over the years on this criterion's very meaning, with the supporters of the concept arguing that the spirit of sport clause never purported to be an absolute definition and that it is merely a characterisation of sport.⁴⁶ Notwithstanding its many criticisms, the spirit of sport remains, to date, the pillar or 'cornerstone justification'⁴⁷ for anti-doping, having survived several revisions of WADC since its first publication in 2003.

1.1 Enhancement Criterion

Strikingly, enhancement is *not* compulsory for the banning of an intervention as it is only one of the three criteria (alternatively, a substance could be banned if it (a) represents an actual or potential risk to the athlete's health and (b) its use is considered contrary to the spirit of sport), nor is *actual* enhancement required to be shown if this criterion were invoked. Thus, the optional applicability of the enhancement criterion implies that a substance can be included on PLIS even if it does not enhance or if it in fact *diminishes* performance. Cannabinoids, for example, are banned despite evidence

⁴⁶ Michael McNamee, 'The Spirit of Sport and The Medicalisation of Anti-Doping: Empirical and Normative Ethics' (2012) 4(4) Asian Bioethics Review <<https://muse.jhu.edu/article/494969>> accessed 29 April 2022.

⁴⁷ Mojsola and Pascal Borry, 'The Landscape of the "Spirit of Sport"' (2019) 16 Journal of Bioethical Inquiry 444.

suggesting that they diminish athletic performance.⁴⁸ This seems to be a somewhat nonsensical phenomenon — surely, substance intake in the absence of enhancement runs counter to our understanding of what doping is. If the purpose of WADC is to prevent illicit enhancement, or ‘cheating’ through gaining an unfair advantage, the enhancement criterion should be compulsory rather than optional.

Not only is it problematic that the applicability of the enhancement criterion is not compulsory, the formulation of the criterion itself is paradoxically and unduly inclusive. Upon analysing the wording of the criterion, subsections 4.3.1.1 and 4.3.1.2 provide that a substance or method may be prohibited if it has a mere ‘*potential*’ to enhance. This means that a substance or technique *does not* have to be proven to enhance but merely that it has potential to do so. Applying this formulation, potential health risks could include ‘eating three plates of spaghetti’ risking obesity, whilst potential to enhance could include ‘six training sessions per week at the age of ten (a method currently employed in numerous sports)’.⁴⁹ In this manner, WADA has taken a strict, deontological approach which focuses on punishing athletes’ *attempts* at gaining an unfair advantage, regardless of whether that unfair advantage is in fact gained. By favouring deontological arguments, the regulation of doping results in a duty-based regulatory framework in which athletes are strictly punished for the breach of their duties. By focusing on punishing the very *act* of doping, regardless of the athlete’s *actual* intention or resulting enhancement (if any), the parameters for WADA’s exercise of discretion have been set incredibly wide, whereby the power to ban an intervention is triggered by the slightest degree of health risk or enhancement potential. It follows that, although the enhancement criterion should be a purely factual matter underpinned by relevant biomedical evidence,⁵⁰ it can (and too often does) become a subjective judgement call made under no known indicative test by WADA.⁵¹ Even if there are such tests, the

⁴⁸ Marilyn A Huestis, Irene Mazzoni and Olivier Rabin, ‘Cannabis in Sport: anti-doping perspective’ (2011) 41 Sports Medicine 949.

⁴⁹ Olivier de Hon, ‘The Redundancy of the Concept Of ‘Spirit of Sport’ in Discussions on The Prohibited List Of Doping Substances’ (2017) 9 International Journal of Sport Policy and Politics 671.

⁵⁰ Sigmund Loland and Hans Hoppeler, ‘Justifying anti-doping: The fair opportunity principle and the biology of performance enhancement’ (2012) 12(4) European Journal of Sport Science 347.

⁵¹ Jonathan Pugh and Christopher Pugh, ‘Neurostimulation, doping, and the spirit of sport’ (2021) 14(2) Neuroethics 141.

reasoning behind their application would not be documented nor available for inspection, and often ‘the minutes of the meetings where [such] decisions [are] made’ are not recorded.⁵² At times, ‘implicit criteria’⁵³ appear to be at work when the contents of PLIS are discussed. Two such examples cited in the literature are the considerations associated with the need to protect athletes themselves from harm, and the preservation of their image as role models to young admirers.⁵⁴

A further way in which the enhancement criterion is excessively expansive is that it captures substances that are not by themselves enhancing but may be such if combined with other substances, as per footnote 27 of WADC: ‘there may be substances that, when used alone, are not prohibited but which will be prohibited if used in combination with certain other substances’. This clarification seeks to encompass a wider set of potential substances and, without further guidance, generates (at least theoretically) an endless list of substances/methods capable of falling within the ban on account of *possible* interactions. As an extreme example, imagine an instance where there is an intake of water along with protein, which then produces an enhancing effect — it would seem to appear that, under the current framework, water and protein can indeed be theoretically banned under the enhancement criteria if and when combined. If so, this may surely have the effect of capturing ‘ordinary’ substances which an athlete would consider using as standard practice. It would then be left to the ambiguous notion of the spirit of sport (as the other criterion is required to reach the two-thirds threshold for banning) which, in theory, might narrow down the list of potential banned substances/methods. In practice, using the notion of the spirit of sport to qualify for the application has its own deep conceptual flaws that hinder its ability to provide a fair and logical solution. This will be discussed in detail in Section 1.2. Evidently, the example given here is hypothetical, but it seeks to show that the above-mentioned criticisms are not simply semantic: they go to the fundamental contention that these criteria are of little practical guidance.

This expansive approach is not only indicative of WADA’s lack of clear guidance on enhancement but can be perceived as an intentional ambiguity on WADA’s part to fulfil its desire for flexibility, which is in line with its

⁵² de Hon (n 49) 671.

⁵³ *ibid* 671.

⁵⁴ *ibid* 672.

pre-emptive approach to *police* the duty not to dope, and its quest 'to get rid of all sorts of vices in the world of sports'.⁵⁵ The pivotal words in WADC are that WADA shall 'consider' in Article 4.3.1, and the word 'potential' in Article 4.3.1.2. Together, they give WADA immense discretion to introduce bans on doping despite the lack of evidence of *actual* enhancement. Indeed, a 2019 study revealed that '[o]nly 5 out of the 23 substance classes show evidence of having the ability to enhance actual sports performance'⁵⁶ and a further 2021 study by Bezuglov et al. revealed that four banned substances on PLIS (meldonium, trimetazidine, xenon and cobalt) bear no evidence of potential performance enhancement at all⁵⁷ — conjuring an uneasy suspicion that WADA may have actually abused the leeway afforded to it with the 'potential to enhance' criterion. The irony of this is that such vague criteria are detrimental to the integrity of sport — the very stated aim of anti-doping programs prominently mentioned in the Preamble of WADC. The lack of guidance opens the decisions on the inclusion of substances or methods to subjective determination, potentially resulting in inconsistencies and biased conclusions.

Although establishing the athlete's state of mind is irrelevant to WADA when considering whether the enhancement criterion has been fulfilled (for which there exists no relevant test), Engelberg et al. note that, in reality, 'many cases of doping feature inadvertent (or accidental) use, where no 'motive' for doping actually exists'.⁵⁸ In many instances, this has been because of carelessness (but, importantly, not because of intention to enhance) on the part of the athlete as to the composition of the drug purchased, as over-the-counter drugs are easily accessible and can contain a variety of banned substances, including prohibited stimulants such as cathine, ephedrine and pseudoephedrine.⁵⁹ The

⁵⁵ de Hon (n 49) 672.

⁵⁶ Jules A A C Heuberger and Adam F Cohen, 'Review Of WADA Prohibited Substances: Limited Evidence for Performance-Enhancing Effects' (2019) 49(4) Sports Medicine 525.

⁵⁷ Eduard Bezuglov and others, 'The Inclusion in WADA Prohibited List is not Always Supported by Scientific Evidence: A Narrative Review' (2021) 12(2) Asian Journal of Sports Medicine 1.

⁵⁸ Terry Engelberg, Stephen Moston and James Skinner, 'The final frontier of anti-doping: A study of athletes who have committed doping violations' (2015) 18(2) Sport Management Review 268, 269.

⁵⁹ David Mottram, 'Aspetar Sports Medicine Journal - Inadvertent Use of Drugs in Sport' (*Aspetar.com*) <<https://www.aspetar.com/journal/viewarticle.aspx?id=217>> accessed 26 April 2022.

riposte from WADA may well be that the PLIS publication is transparent and it is a duty of every athlete to stay vigilant. However, for many athletes, a dedicated backroom team of medical staff to scrutinise their diets and medicine intake is not feasible. Indeed, many athletes do not even control their own diets.⁶⁰ Furthermore, representing one's own country at international elite events such as the Olympics does not simply translate into access to sufficient funding to ensure compliance. For example, Canadian skeleton racer Mirela Rahneva explained that, even though she is part of the Canadian Assistance Program, she had to meet her financial needs through her own money and through setting up a GoFundMe page, with per-season expenses totalling over \$26,000.⁶¹ Heavily burdened by these financial constraints, it would not be surprising if athletes failed to keep up with the ever-changing PLIS list. This problem is further exacerbated when athletes are not representing their countries (for example, in Grand Slam tennis matches) and therefore do not receive government funding. This means that, in such situations, athletes are left financially unsupported and are burdened with ensuring PLIS compliance despite not having any medical expertise. Even athletes who can afford backroom teams of support may fall victim to the constant changes: Maria Sharapova was banned from professional tennis in March 2016 after testing positive for meldonium, a drug that was only added to PLIS two months prior to her taking it, but which she had been taking for the past ten years to deal with her 'irregular heartbeat and a history of diabetes in her family'.⁶² If even global superstars like Sharapova cannot keep up with WADA's constant changes, what hope do less well-funded athletes have?

The enhancement criterion is therefore one that is all-encompassing and simultaneously vague, resulting in some odd inclusions on PLIS, which is detrimental to the integrity of sport as innocent athletes are found guilty of enhancement despite lacking the requisite intention. WADA's possible retort

⁶⁰ 'Athletics: Ostapchuk's coach admits drugging her food' (*The New Zealand Herald*, 2012) <<https://www.nzherald.co.nz/sport/athletics-ostapchuks-coach-admits-drugging-her-food/UCFX3LVX7HMIHJQDOY23LXICY/>> accessed 27 September 2022.

⁶¹ Olivier Neven, 'Skeleton racers blast Canada's lack of Olympic funding: "We saw \$0"' (*Yahoo Sports Canada*, 20 February 2022) <<https://uk.sports.yahoo.com/news/beijing-olympics-canadian-skeleton-racers-blast-canada-lack-of-funding-0-dollars-005753264.html>> accessed 25 November 2022.

⁶² Les Carpenter, 'Maria Sharapova Provisionally Banned From Tennis After Revealing Failed Drugs Test' (*the Guardian*) <<https://www.theguardian.com/sport/2016/mar/07/maria-sharapova-failed-drugs-test-australian-open-2016-tennis>> accessed 26 April 2022.

here may be that the expansive nature of the enhancement criterion is mitigated by the third criteria: the spirit of sport. This will be discussed in Section 1.2 below. However, as we shall see further, this criterion is plagued with even more issues of its own.

1.2 The Spirit of Sport Criterion

Like its enhancement criterion counterpart, the spirit of sport is also not compulsory. This means that a substance/method can be included in PLIS based on the other two criteria, despite not being deemed to have violated the spirit of sport. However, in practice, it is difficult to see how the enhancement criterion is separable from that of the spirit of sport. For example, when an athlete takes a performance enhancing drug, it could be argued that they have fallen afoul of the ‘honesty’ value of the spirit of sport (see below), too. Indeed, WADA itself notes that ‘doping is fundamentally contrary to the spirit of sport’.⁶³ This suggests that, where there is doping, there must be a violation of the spirit of sport and vice versa. As such, it appears odd that the spirit of sport criterion is *optional*. If it is indeed possible that a substance/method can enhance but not violate the spirit of sport, it is inconceivable that this criterion should be the ‘fundamental rationale’⁶⁴ against doping. If, on the other hand, as WADA suggests, *any* doping violates the spirit of sport, then the concept is itself redundant because it would be ‘superfluous to repeat the same argument twice’.⁶⁵

What then exactly is the spirit of sport? WADC states that it is ‘the celebration of the human spirit, body and mind. It is the essence of Olympism’⁶⁶ and ‘Olympism’ is defined in IOC’s Olympic Charter as:

a philosophy of life, exalting and combining in a balanced whole the qualities of body, will and mind ... seeks to create a way of life based on the joy of effort, the educational value

⁶³ WADC (n 45) 13.

⁶⁴ *ibid* 13.

⁶⁵ de Hon (n 49) 673.

⁶⁶ WADC (n 45) 13.

of good example, social responsibility and respect for universal fundamental ethical principles.⁶⁷

The spirit of sport is also all the values we find ‘in and through sports’,⁶⁸ and 12 such values are listed in WADC and reproduced below:⁶⁹

- Health
- Ethics, fair play and honesty
- Athletes’ rights as set forth in the Code
- Excellence in performance
- Character and education
- Fun and joy
- Teamwork
- Dedication and commitment
- Respect for rules and laws
- Respect for self and other participants
- Courage
- Community and solidarity

What is immediately apparent from the lexicon used is that the spirit of sport is less a definition than a characterisation or a declaration of assorted aspirational principles. The existence of this list can be traced back to various scandals dating back to the 1960s, resulting in the creation of the spirit of sport ‘ideal’, one which witnessed the transformation of the original health rationale for anti-doping into isolated attacks on morality of athletes.⁷⁰ In attempting to balance the interests of various stakeholders, an open-ended definition was adopted to appease the public whilst maintaining its own legitimacy.⁷¹ The resulting regime was one which lacked sustainable ethical viewpoints. As such, WADA has been left with a legacy of contradictions, which have been carried

⁶⁷ International Olympic Committee, ‘Olympic Charter’ (*Stillmed.olympics.com*, 8 August 2021) 8

<https://stillmed.olympics.com/media/Document%20Library/OlympicOrg/General/EN-Olympic-Charter.pdf?_ga=2.71261845.469085637.1651305364-1539759114.1651305364> accessed 30 April 2022.

⁶⁸ WADC (n 45) 13

⁶⁹ *ibid* 13.

⁷⁰ Ian Ritchie, ‘The construction of a policy: The World Anti-Doping Code’s ‘spirit of sport’ Clause’ (2013) 2 *Performance Enhancement & Health* 194, 195.

⁷¹ *ibid* 199.

forward to the present. These principles, problematically, fail to provide the much-needed definitional certainty to an otherwise mystifying notion of the spirit of sport. Unsurprisingly, it has been dismissed as ‘simply utopian’,⁷² a ‘disparate list’⁷³ of little worth, offering only ‘an incomplete, unsystematic and unstructured account of key values in ethical sport’.⁷⁴ Insofar as Olympism is concerned, the concept has been criticised as seriously outdated — a futile embrace of a deeply ‘socially and historically conditioned ideal’ created by the myth of the marathon in 490 BC.⁷⁵

These criticisms are not groundless: these principles are vague and far from helpful in establishing what the spirit of sport is. Indeed, these values seem applicable even outside the world of sports. For example, a person holding themselves out as a lawyer without passing the relevant exams could be considered to have a lack of respect for the ‘rules and laws’ and thereby would seem to be in violation of the spirit of sport criterion. The point here is that such a fundamental list is not tailored to sports at all. This unfortunate characteristic stems from the historic origins of the ‘spirit of sport’: the emergence of these values and the idealism of Olympism reflects the immense pressure on the Code Project Team of WADA at the turn of this century when WADC was created. Faced with scandals involving state-sponsored doping in the 1980s⁷⁶ and the ‘Festina’ affair⁷⁷ about systematic doping at the Tour de France at the end of the 1990s, much was expected from the creation of WADA. WADC was needed not only to overcome the enduring problem of defining doping, but was also expected to restore the tarnished image of all of sport. Inevitably, then, the lexicon used was filled with positive and aspirational

⁷² de Hon (n 49) 671.

⁷³ Sigmund Loland and Michael J. McNamee, ‘The “spirit of sport”, WADA’s Code Review, and the search for an overlapping consensus’ (2019) 11(2) *International Journal of Sport Policy and Politics* 325, 333.

⁷⁴ Pugh and Pugh (n 51) 143.

⁷⁵ Loland and McNamee (n 73).

⁷⁶ ‘The State-sponsored Doping program’ S8 EP1: Doping for Gold (*pbs.org*, 13 June, 2011) <<https://www.pbs.org/wnet/secrets/the-state-sponsored-doping-program/52/>> accessed 25 November 2022.

⁷⁷ Jeremy Whittle, ‘Twenty Years on the Festina Affair Casts Shadow Over the Tour De France’ (*the Guardian*, 3 July 2018) <<https://www.theguardian.com/sport/2018/jul/03/tour-de-france-festina-affair>> accessed 26 April 2022.

language with the goal of not just reducing the prevalence of doping, but also to provide for a central ‘educational platform’ to youngsters.⁷⁸

Despite these admirable goals, almost all 12 values have come under attack in one form or another. For example, the value of ‘honesty’ has been labelled as unhelpful since an athlete using a drug not yet prohibited should not be regarded as dishonest.⁷⁹ By the same logic, a lack of ‘respect for rules and laws’ could not ‘be logically invoked to justify a ban ... as they depend on the prior existence of such a ban’.⁸⁰ Paradoxically, it might even be argued that doping can help realise some of these values. For example, sacrificing one’s health by taking on the risk of a new experimental drug to ‘bring [one’s] athletic performance to a new level’⁸¹ may be construed as ‘a sign of courage’,⁸² and the use of doping might be seen as the pursuit of ‘excellence in performance’ and a symbol of ‘dedication and commitment’⁸³. McNamee has attempted to offer a defence, arguing that conceptual vagueness is intrinsic to all natural languages, and cited that, in football terms, we are likely to agree to what constitutes ‘excessive force’ when applying rules prohibiting such force.⁸⁴ However, such an argument can hardly be convincing given the central role that the spirit of sport plays in underpinning the *entire* regulatory framework of WADA, where ambiguity only serves to amplify uncertainty and the risks of arbitrary discretion. This argument is not to suggest that WADA should adopt a strict and rigid definition of the spirit of sport; instead, as Section III suggests, a decentralised model allowing for greater input from specific sporting bodies provides for a much more realistic manner of regulation.

Another problem with this concept is that it assumes there is a ‘spirit’ that is attainable and tangible. It takes a teleological view of sports in that there is some purpose to playing and competing. Indeed, Sandel writes that ‘[a]rguments about the ethics of enhancement are always, at least in part,

⁷⁸ Ritchie (n 70).

⁷⁹ Vincent Geeraets, ‘Ideology, Doping and The Spirit of Sport’ (2017) 12(3) Sport, Ethics and Philosophy 255.

⁸⁰ Elizabeth Shaw, ‘Neurodoping in Chess to Enhance Mental Stamina’ (2021) 14(2) Neuroethics 220.

⁸¹ Erler (n 10) 200.

⁸² Geeraets (n 79) 258.

⁸³ WADC (n 45) 13.

⁸⁴ McNamee (n 46).

arguments about the telos, or point, of the sport in question, and the virtues relevant to the game'.⁸⁵ But the accuracy of this assertion is doubtful in an increasingly commercialised world of sports, where incentives to win far exceed the potential consequences of violating the spirit of sports; as such, many athletes no longer internalise the idea of the spirit of sport. For many athletes, the telos of sports is no longer 'the celebration of the human spirit', if it ever was.

Per Forbes, footballer Lionel Messi earned \$97 million in salary in 2021 alone, with a further \$33 million in off-pitch earnings, whereas an average Premier League footballer earned \$85,000 in salary per month, or \$1.02 million per year.⁸⁶ By comparison, the annual median UK salary in 2021, calculated using the data gathered by the Office for National Statistics, was £25,990.⁸⁷ These figures show that there is a huge monetary incentive for athletes to use any means possible to try and reach their athletic potential, not necessarily for the fulfilment of some idealistic view of sport, but for monetary gain. Although at significantly lower sums, chess largely operates in the same way, with income derived from prize money, sponsorships, online streaming and, at one extreme end, public corporate finance opportunities such as the listing of chess company Play Magnus on the Oslo stock exchange by the reigning world chess champion Magnus Carlsen in 2013, which had the fortune to blossom from a mere training app into the 'only publicly traded chess company in the world'⁸⁸ to date.

It is not suggested here that the lure of millions of dollars in prize money and sponsorships necessarily means that athletes can no longer

⁸⁵ Michael J Sandel, *The Case Against Perfection: Ethics in the Age of Genetic Engineering* (Harvard University Press 2007) 38.

⁸⁶ Ellis Cashmore, Jamie Cleland and Kevin Dixon, 'Football Is Ready for Openly Gay Players' (*Fair Observer*, 24 February 2021) <<https://www.fairobserver.com/culture/ellis-cashmore-jamie-cleland-kevin-dixon-openly-gay-athletes-football-lgbtq-news-17281/>> accessed 26 April 2022.

⁸⁷ 'Earnings And Hours Worked, All Employees: ASHE Table 1 — 2021 Revised Edition' (*Office for National Statistics*) <<https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/earningsandworkinghours/datasets/allempleeasasetable1>> accessed 26 April 2022.

⁸⁸ Owen Poindexter, 'How World Champion Magnus Carlsen Rode the Chess Boom' (*Front Office Sports*, 30 November 2021) <<https://frontofficesports.com/how-world-champion-magnus-carlsen-rode-the-chess-boom/>> accessed 28 April 2022.

internalise what the spirit of sport entails. However, athletes often unsurprisingly find themselves caught in a prisoner's dilemma where the lure of financial and professional success may outweigh the risk of getting caught and receiving a ban from competition.⁸⁹ Savulescu et al. point out that 'the enormous rewards for the winner, the effectiveness of the drugs, and the low rate of testing all combine to create a cheating "game" that is irresistible to athletes'.⁹⁰ Indeed, this lure to 'cheat' is perhaps strongest amongst the majority of professional athletes who are unlikely to become founders of listed companies like Carlsen and would depend on prize money and sponsorships to maintain a living. Comparatively speaking, achieving success through 'legitimate' means, such as training and equipment, may be more expensive than through illicit doping.

Within this equation also lie the aspirations of sponsors, national bodies and fellow countrymen — sources of pride as well as pressure. Boye et al. explain that, as achievements in sports are closely tied to national pride, temptations to cheat can be stronger than respect for the rules or concerns about the health of athletes.⁹¹ National funding acts as a particular pressure point that accounts for why a national athlete would choose to dope by taking a 'win at all costs' mentality.⁹² Indeed, a direct correlation exists between the number of medals won at the Olympics and the amount of funding. For example, in 2016, Ingold and Williams completed a survey demonstrating the correlation between the level of funding from 1999 to 2015 and the total number of US Olympic medals won by the US from 2004 to 2014. The empirical evidence clearly demonstrated that the most medals are won in highly funded sports, such as

⁸⁹ Julian Savulescu, Bennett Foddy and M Clayton, 'Why We Should Allow Performance Enhancing Drugs in Sport' (2004) 38 *British Journal of Sports Medicine* 666; Jill S Harris, 'Doping on Deck: The Prisoner's Dilemma of Performance-Enhancing Drugs' (2020) *The Economics of Aquatic Sports* 67.

⁹⁰ Savulescu and others (n 89) 667.

⁹¹ Erik Boye and others, 'Doping and Drug Testing' (2017) 18 *EMBO Reports* 351-354.

⁹² See, for example: "'Athletes Have Had Enough': Money, Medals Putting Pressure on Canadian High-Performance Sport Culture | CBC Sports" (*CBC News*, April 20 2022) <<https://www.cbc.ca/sports/olympics/olympics-canada-culture-analysis-1.6423878>> accessed 2 November 2022.

swimming, track and field, skiing and snowboarding.⁹³ The less funded sports, such as gymnastics and ice hockey, brought home the fewest medals.

Thus, the claim that sports are a ‘celebration of the human spirit, body and mind’ is divorced from the commercial reality that high performance is a luxury that not all athletes can afford. Sports, especially for those in less privileged countries, are often played out of necessity. Oxana Chusovitina, a decorated Olympian whose son was diagnosed with acute lymphocytic leukaemia, relocated to Germany from her home country of Uzbekistan which had no specialist oncology facilities and where her family lacked health insurance.⁹⁴ At the time, she noted: ‘If I don’t compete, then my son won’t live, it’s as simple as that’.⁹⁵ Competition is often a means of salvation, not a celebration of spirit. Even in wealthy countries, such as the US, sports are seen as a route out of poverty. LeBron James, arguably the main star of modern basketball, admitted that basketball was a ‘vehicle’ he used to ‘get to where [he] want[ed] to go’.⁹⁶ Reflecting on his success, he stated: ‘[w]ow, you know, we didn’t let the game of basketball use us, we used it’.⁹⁷ Strikingly, basketball was seen as a means to an end, and not, as WADA would suggest, an end in itself. These stories, and countless others, suggest a much more pessimistic view of sports. Rather than being a ‘celebration’, the calculation is far more utilitarian — athletes sacrifice their bodies: in the case of Chusovitina, she had to compete well past the usual age of retirement for the survival of her son.

⁹³ David Ingold and Eben Novy-Williams, ‘Money for Medals: Inside the Performance-Driven Funding Of U.S. Olympic Teams’ (*Bloomberg*, 5 August 2016) <<https://www.bloomberg.com/graphics/2016-olympics-usoc-return-on-investment/>> accessed 30 April 2022.

⁹⁴ Eric Kirschbaum, ‘Child’s Cancer care brough gymnast to Germany’ (*Reuters*, 6 June 2008) <https://www.reuters.com/article/us-olympics-gymnastics-chusovitina-idUSL2281235320080606>> accessed 28 October 2022.

⁹⁵ Emma John, ‘Oksana Chusovitina’s Epic Career Provides Many Reasons to Be Tearful’ (*The Guardian*, 4 November 2018) <<https://www.theguardian.com/sport/2018/nov/04/gymnast-oksana-chusovitina-saga-brings-tear-to-eye?bcmt=1>> accessed 28 April 2022.

⁹⁶ Tony Manfred, ‘Lebron James Really Did Grow Up in Inner City Poverty’ (*Business Insider*, 21 June 2013) <<https://www.businessinsider.com/lebron-james-life-story-2013-6>> accessed 28 April 2022.

⁹⁷ *ibid.*

In sum, neither Olympism nor the list of values are of much ‘operative utility’.⁹⁸ It is therefore most surprising that, despite the vagueness, it has remained intact since its inception in 2003, and survived four rounds of WADC revisions. Although no changes were made in the first two rounds (the 2009 and 2015 versions), various changes were introduced during the third round (the 2019 version), which were endorsed unchanged in the fourth and the latest 2021 version, and compiled, according to WADA, ‘under the watchful eye of WADA’s Legal Department, which coordinated the work ... to ensur[e] consistency and harmonization’.⁹⁹ Despite these changes, harmonisation is still difficult to discern, with the additions and re-orderings failing to fundamentally clarify *what* exactly the ‘spirit of sport’ means, or what the ‘underlying logic of [the chain of values] inclusion or hierarchy or coherence’ is.¹⁰⁰ In turn, what the amendments did signal is the rather unfortunate message that the spirit of sport is conceptually, and in its updated form, *alive and well*.

The spirit of sport is a tall order and has been proffered as a criterion, an underlying ‘philosophical definition’¹⁰¹ and a ‘sound ethical justification for banning drugs and methods’¹⁰² *all at once*. Despite this lofty goal, on the eve of its 20th birthday, widespread doping use and contraventions continue to exist. The Russian Federation was banned from competing in major competitions including the Olympics,¹⁰³ only to find that ‘neutral’ Russian athletes were allowed to compete under its ‘replacement’ Russian Olympic Committee flag. The latest drama at the Beijing Winter Olympics 2022 involved the 15-year-old ice-skating sensation Kamila Valieva, who was allowed to compete *despite*, on *any* account of *any* interpretation of the spirit of sport, testing positive for trimetazidine. The reason cited was that she was ‘a Protected Person’.¹⁰⁴ In 2021,

⁹⁸ Loland and McNamee (n 73) 333.

⁹⁹ ‘WADA Code Review’ (WADA, 2022)

<<https://www.wada-ama.org/en/what-we-do/world-anti-doping-code/code-review>> accessed 30 April 2022.

¹⁰⁰ Loland and McNamee (n 73) 333.

¹⁰¹ Ritchie (n 70) 199.

¹⁰² *ibid* 198.

¹⁰³ Billy Perrigo, ‘Russia Banned from Olympics, World Cup, For State-Sponsored Doping’ (*Time*, 9 December 2019)

<<https://time.com/5746344/russia-banned-olympics-2019/>> accessed 28 April 2022.

¹⁰⁴ ‘Beijing 2022: The ITA Informs on Figure Skater Kamila Valieva’ (*International Testing Agency*, 11 February 2022)

it was estimated that between 10% and 40% of athletes might have cheated at the 2020 Tokyo Olympics.¹⁰⁵ Brand et al. note that ‘the percentage of doping cases which remain unrevealed is thought to be considerably higher’¹⁰⁶ and Sottas et al. suggest that blood doping was present in 14% of elite track and field athletes on average.¹⁰⁷ These astonishing statistics show that, despite WADA’s best efforts, the ‘gold medal culture’ prevails as the *de facto* spirit of sport.

1.3 Physical/Mind Sports Divide — Over-Inclusive

Taking a step back from the flawed two-thirds criteria, another observation can be made of WADA’s current regulation, namely that it fails to distinguish between physical and mind sports.

PLIS, compiled in accordance with the criteria in Article 4.3 of WADC, is currently sorted into substances prohibited only in competition (a smaller category disallowed 12 hours before competition), substances prohibited both in and out of competition and substances applicable to certain precision sports only (mainly beta-blockers, which reduce tremors).¹⁰⁸ This lack of distinction between physical and mind sports is unfortunate and shall be explored in more detail in Section 2.1 below. This article proposes that any future change to regulation should be firmly grounded on such a distinction if sustainable policies are to be devised — muscle building via anabolic steroids would not improve Carlsen’s chess game; likewise, improved mental stamina through tDCS would not assist Usain Bolt’s 100m time..

<<https://ita.sport/news/beijing-2022-the-ita-informs-on-figure-skater-kamila-valieva/>> accessed 29 April 2022.

¹⁰⁵ ‘Sport Is Still Rife with Doping’ (*The Economist*, 14 July 2021)

<<https://www.economist.com/science-and-technology/2021/07/14/sport-is-still-rife-with-doping>> accessed 28 April 2022.

¹⁰⁶ Ralf Brand, Wanja Wolff and Detlef Thieme, ‘Using Response-Time Latencies to Measure Athletes’ Doping Attitudes: The Brief Implicit Attitude Test Identifies Substance Abuse in Bodybuilders’ (2014) 9 *Substance Abuse Treatment, Prevention, and Policy*.

¹⁰⁷ Pierre Sottas and others, ‘Prevalence of Blood Doping in Samples Collected from Elite Track and Field Athletes’ (2011) 57 *Clinical Chemistry* 762.

¹⁰⁸ ‘WADA The Prohibited List’ (*WADA*)

<<https://www.wada-ama.org/en/prohibited-list/>> accessed 29 April 2022.

Being appreciative of the physical/cognitive divide is not to suggest that cognition is irrelevant in physical sports or that physical condition is irrelevant in mind sports. An example of the former is gymnastics: a gymnast cannot pin the perfect landing on physical aptitude alone — the gymnast must perfectly sequence and coordinate such moves through brain processing or so-called ‘perceptual-cognitive skills that influence the anticipation and decision-making processes’,¹⁰⁹ which sets champions apart. In other words, ‘game intelligence’.¹¹⁰ Conversely, although grandmasters can be stationed at the chess table for hours on end, chess is very much a physical activity¹¹¹ with players in ‘faster games’, for example, the World Blitz Championships (where players are only allocated five minutes each) moving ‘at least 1,000 chess [pieces] in addition to 1,000 activation[s] of the chess clock by finger pressure’.¹¹²

Instead, the commitment to the physical/cognitive divide should be construed fundamentally as an appreciation of the differences in the *kind* of sport in question. For sports with a strong mental element, cognitive enhancement will be especially relevant. This includes sports that require creativity and problem-solving skills, such as chess,¹¹³ bridge and Esports.¹¹⁴ On the other hand, for sports with a more mechanical element, physical enhancement will be of special relevance. For example, the use of anabolic steroids which include natural androgens like testosterone has the effect of increasing protein within cells, especially in skeletal muscles. Ultimately, it is a matter of *degree* as to which element — cognitive or physical — a particular sport is oriented towards and, whilst there inevitably is some overlap between physical and cognitive enhancement (e.g., cyclists can benefit from both EPO

¹⁰⁹ Filipe Casanova and others., ‘Expertise and perceptual-cognitive performance in soccer: A review perícia e rendimento perceptivo-cognitivo no futebol’ (2009) *Uma Revisão da Literatura* 120.

¹¹⁰ *ibid* 116.

¹¹¹ Sighart Golf, ‘Biochemistry and Psychology of Chess and Classical Physical Exercise: Concurring or Conflicting Evidence?’ (2015) 5 *Journal of Sports Medicine & Doping Studies* 3.

¹¹² *ibid*.

¹¹³ Shaw (n 80) 223.

¹¹⁴ Sergio Machado and others, ‘Could Tdcs Be a Potential Performance-Enhancing Tool for Acute Neurocognitive Modulation in Esports? A Perspective Review’ (2021) 18 *International Journal of Environmental Research and Public Health*.

and tDCS for enhanced endurance),¹¹⁵ the key is that some sports stand to benefit more from one form of enhancement than another.

Another reason to commit to this divide is that it enables the appreciation of the differences in establishing a link between doping and enhanced performance. Whilst it is often easier to establish this link in the physical domain, the same cannot be said of cognitive enhancement. For example, in the case of EPO, the link between oxygen carrying capacity and endurance performance is intimately related, but the same link does not hold for cognitive enhancement where it is difficult to pin down one specific causal link — is it the result of a change in neuroplasticity, a chemical reaction, or both? In other words, the ‘mechanism’ through which enhancement occurs is not always apparent in relation to cognition.¹¹⁶ It is of little surprise, therefore, that sports which are more mechanical in nature exhibit higher doping figures, such as athletics and weightlifting (as shown in *Figure 2*),¹¹⁷ accounting for 68% out of 57 Olympic medals stripped so far (from 1968-2012) at 42% and 26% respectively.¹¹⁸

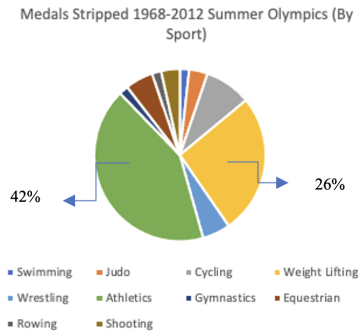


Figure 2: He and Murray (2016)

¹¹⁵ Petersen and Lippert-Rasmussen (n 14).

¹¹⁶ Pugh and Pugh (n 51) 149.

¹¹⁷ The pie chart was compiled with figures taken from Nicole He, K K Rebecca Lai and Paul Murray, ‘Athletes Who Were Denied Their Olympic Medal Moments Because Others Were Doping’ (NY Times, 14 August 2016)

<<https://www.nytimes.com/interactive/2016/08/14/sports/olympics-medal-doping.html>> accessed 28 April 2022.

¹¹⁸ Pugh and Pugh (n 51).

Although individual athletes may dope for a variety of reasons, figures do go some way to support the claim that there must be a connection between the *type* of sport and the *efficacy* of doping — with the more physical or mechanical sports deriving the greatest benefit (which, therefore, explains the higher number of dopers). It may be argued that *how* the intervention occurs is entirely irrelevant since WADA takes a strict liability approach that captures every substance that potentially enhances irrespective of *how* it enhances. This lack of appreciation of nuance supports the claim made in this article that the anti-doping regulations, as currently constructed, are intrinsically flawed and accounts for WADA's inertia to make necessary changes. This will be further discussed in Section II.

This intentional or unintentional tendency to draw misleading parallels between mind sports and physical sports will only be exacerbated by the growth in NIBS technologies, as the following Section explores.

1.4 Cognitive Enhancement — Under-Inclusive and Under-Explored

Currently, PLIS sets out prohibited substances, including nootropics (S0-S9), and methods such as blood/gene manipulation (M1-M3). However, it does not *yet* prohibit NIBS or other forms of cognitive enhancement methods. This is not to say, however, that WADA does not regulate cognitive enhancement at all. To be clear, PLIS does include substances believed to enhance cognition. Oft-cited substances and their perceived areas of cognitive enhancement, together with their respective status, are set out in a non-exhaustive table below. These substances can otherwise be classified as 'nootropics', or pharmacological substances that have the effect of cognitive enhancement. Although there is some controversy as to the regulation of caffeine, especially after an increase in uptake by athletes after its removal from PLIS,¹¹⁹ the general consensus around nootropics is that there is an obvious enhancing effect that justifies their regulation.

¹¹⁹ Neil Chester and Nick Wojek, 'Caffeine Consumption Amongst British Athletes Following Changes to the 2004 WADA Prohibited List' (2008) 29 *International Journal of Sports Medicine*.

	Substance	Cognitive enhancing effect	PLIS listed?
1	Amphetamine	Improve consolidation of information, leading to improved recall ¹²⁰	Yes (S6 — in competition)
2	Caffeine	Increase alertness, attention and performance, ¹²¹ improve endurance	No — expressly excluded from S6 class
3	Methylphenidate	Improve performance in novel tasks, reduce planning latency in complex tasks ¹²²	Yes (S6 — in competition)
4	Modafinil	Improve reaction time, logical reasoning, problem solving ¹²³	Yes (S6 — in competition)

Although this table shows that there is some degree of consideration of the effects of cognitive enhancement by WADA, two realities render the current regulation ineffective. First, this article contends that, although WADA does ban substances capable of cognitive enhancement, it does not do so through a systematic, considered and principled approach. Since WADA's inception, much attention has been paid to prohibit physical enhancement — to be precise, physical doping in physical sports — commonly through the use of drugs such as steroids and EPO with headline-grabbing cases. Examples include

¹²⁰ Kara Simone Bagot and Yifrah Kaminer, 'Efficacy of Stimulants for Cognitive Enhancement in Non-Attention Deficit Hyperactivity Disorder Youth: A Systematic Review' (2014) 109 *Addiction*.

¹²¹ Bernadette Marriott, 'Food Components to Enhance Performance: An Evaluation of Potential Performance-Enhancing Food Components for Operational Rations' [1994] Institute of Medicine (US) Committee on Military Nutrition Research.

¹²² *ibid* 315.

¹²³ *ibid* 315.

Lance Armstrong,¹²⁴ Tyson Gay¹²⁵ and, more recently, the UK athletics 4x100m relay team being stripped of their silver medal at the Tokyo 2020 Olympics.¹²⁶ Conversely, much less attention has been paid to cognitive enhancement, in both physical and mind sports, especially in the latter. To date, no chess-player has tested positive for a violation of PLIS. Generally speaking, the impact of cognitive enhancement feeds into all sports, although benefits are perceived to be more pronounced in mind sports and appropriate WADA policies can only be formulated when these nuances are appreciated and well understood.

Secondly, given that WADA currently regulates the use of some substances that may cognitively enhance (for example, Ritalin)¹²⁷ but not all (for example, NIBS technologies that promise to offer multiple cognitive enhancing benefits), there exists a glaring (and growing) regulatory gap for emerging cognitive-enhancing technologies, such as tDCS, tRNS, tACS and EEG. In other words, WADA does not currently regulate *all* the ways in which an athlete can be cognitively enhanced, which means there is a potential regulatory vacuum, which is far from satisfactory. Mindful of this regulatory vacuum, along with the underexplored area of cognitive enhancement, this article will seek to address this lacuna with the help of tDCS and chess, which shall now be turned to.

PART B: A CASE FOR NEURO-DOPING?

This article will now turn to analyse the use of tDCS within WADA's framework, and whether it would be considered a prohibited technique. The

¹²⁴ 'Lance Armstrong 'Stripped' Of 'Tour de France titles and banned' (BBC Sport, 24 August 2012) <<https://www.bbc.com/sport/cycling/19369375>> accessed 28 April 2022.

¹²⁵ Gene Cherry, 'U.S. Sprinter Tyson Gay Banned One Year for Doping' (*Reuters*, 2 May 2014) <<https://www.reuters.com/article/us-athletics-gay-doping-idUSKBN0D11ED20140502>> accessed 21 April 2022.

¹²⁶ Ben Rumsby, 'Team GB Stripped of Olympic Men's 4x100m Relay Silver Medal for Doping Violation' (*The Telegraph*, 18 February 2022) <<https://www.telegraph.co.uk/olympics/2022/02/18/team-gb-stripped-olympic-mens-4x100m-relay-silver-medal-doping/>> accessed 28 April 2022.

¹²⁷ 'WADA Therapeutic Use Exemptions' (*WADA*) <<https://www.wada-ama.org/en/athletes-support-personnel/therapeutic-use-exemption-s-tues>> accessed 29 April 2022.

following will establish that under the two-thirds criteria, tDCS could in theory be banned by WADA.

1.5 Can tDCS be banned under current WADA regulations?

a) *Health criterion*

There is general ‘scientific consensus so far ... that tDCS poses no major health risks’.¹²⁸ Indeed, there has been growing recognition of therapeutic use of this technology in areas such as MDD.¹²⁹ Therefore, it likely does not fulfil the health criterion.

b) *Enhancement criterion*

There are no known direct studies on the effect of tDCS on chess, but two areas of cognitive enhancement are particularly relevant and will be singled out to establish a correlation with chess performance: memory and mental fatigue. The prevailing view in the literature as to the importance of memory in chess is illustrated by the Chunking hypothesis, developed initially by de Groot in a chess experiment.¹³⁰ The hypothesis purports that ‘the differences in the ability of chess players at different skill levels to copy and to recall positions are attributable to the experts’ storage of thousands of chunks ... in long-term memory [LTM]’.¹³¹ With that, a logical conclusion can be drawn: to the extent that memory recall is adversely impacted, so is LTM and, hence, chess performance. This point may be of special significance for high-level competition where recollection of the opponent’s specific preparatory materials, including a vast repertoire of openings, variations and tactics, is invaluable. In

¹²⁸ Shaw (n 80) 219; see also Laura Sophie Imperatori, Luke Milbourn and Mirko Daniel Garasic, ‘Would the Use of Safe, Cost-Effective tDCS Tackle Rather Than Cause Unfairness in Sports?’ (2018) 2 *Journal of Cognitive Enhancement* 377-387.

¹²⁹ NHS, ‘Treatment - Clinical Depression’ (NHS) <https://www.nhs.uk/mental-health/conditions/clinical-depression/treatment/> accessed November 26 2022.

¹³⁰ Adrianus Dingeman de Groot, *Thought and Choice in Chess* (Amsterdam University Press 2008).

¹³¹ Fernand Gobet and Herbert A Simon, ‘Expert Chess Memory: Revisiting the Chunking Hypothesis’ (1998) 6 *Memory*.

relation to mental fatigue, a study was conducted by Salehi et al.¹³² in 2021 into the effects of tDCS on the dorsolateral prefrontal cortex regarding mental fatigue and physical performance in professional swimmers. This study concluded that anodal tDCS ‘reduces adverse effects of mental fatigue’ in swimming performance. Although these conclusions were made in the context of swimming, reduction of the adverse effects of mental fatigue is not sport-limited, and Salehi et al.’s conclusion can be extrapolated to apply to Ghorbani et al.’s 2020 study, which found that exposing chess players to mental fatigue ‘impaired pattern recall performance of expert chess-players’.¹³³

The link that can be drawn between Ghorbani’s chess study and Salehi’s tDCS study is this:

- P1: tDCS can cause a reduced sensation of mental fatigue in swimmers and in elite chess players alike.
- P2: mental fatigue impacts pattern recall.
- P3: affected ability to recall impacts on LTM and, hence, on chess performance.
- C: improved mental stamina generated by tDCS (as hypothesised) will therefore enable chess players to tap into the chunks of their LTM for longer, thus allowing better recall and a delay in any drop in performance.

Alternatively put, tDCS may create a delay in playing an incorrect move out of fatigue or perception of fatigue (assuming a wrong move is unavoidable at some point) which, *in theory*, increases the player’s chances of winning by avoiding, or delaying, the need to make a series of subsequent moves to ‘save’ the predicament generated by an earlier mistake. As such, a delayed onset of fatigue, or perception of fatigue, may alter the course of the rest of the game (and, possibly, the ultimate result). In this manner, chess players may be cognitively enhanced by tDCS. Bearing in mind that WADA bans what may *potentially* enhance, it is demonstrated here that, at the very least, tDCS does have

¹³² Elahe Nikooharf Salehi and others, ‘Transcranial Direct Current Stimulation Reduces the Negative Impact of Mental Fatigue on Swimming Performance’ (2021) 54 Journal of Motor Behavior 327-336.

¹³³ Saeed Ghorbani, ‘Effects of Mental Fatigue on Memory Function of Expert Chess Players’ Iranian Journal of Learning and Memory 2020 2(8).

this potential to enhance in chess. With this in mind, the enhancement criterion would likely be considered fulfilled.

c) *Spirit of sport criterion*

In Section 1.2, the shortcomings of the spirit of sport were identified, with the vagueness inherent in its definition especially questioned. Indeed, this lack of a precise definition of the spirit of sport means that it is difficult to apply this criterion to tDCS. For the purposes of this article, ethical grounds underpinning the concept will be examined in determining if the spirit of sport has been violated. To this end, three of the most recent key amendments to the Preamble (suggested to be indicative of WADA's latest policy direction) will be examined:

- **Hard Work:** The word 'ethical' has been inserted before 'pursuit of excellence' to read: 'this intrinsic value is often referred to as the "spirit of sport": the *ethical* pursuit of human excellence through the dedicated perfection of each Athlete's natural talents'.¹³⁴ This addition laid down an open-textured requirement of being 'ethical', for which there is no universal consensus as to its meaning. However, since this was stated to be achieved 'through dedicated perfection', the virtues of hard work and commitment can be implied. This ethical justification is premised on the basis that enhancement reduces the need for effort, thus undermining the 'essence' of a sport. The contention is that doping corrodes character and 'reduces the suffering, pain, hard work, and effort that are essential for real and worthy achievements'.¹³⁵ Effort and perseverance are admired because they are morally valuable, under the individual's control,¹³⁶ and reflect transferable skills benefiting society.¹³⁷ It follows that doping must be deplored, as it reduces the

¹³⁴ WADC (n 45) (emphasis added).

¹³⁵ Maartje Schermer, 'Enhancements, Easy Shortcuts, and the Richness of Human Activities' (2008) 22(7) *Bioethics*.

¹³⁶ John Martin Fischer and Mark Ravizza, *Responsibility and Control* (Cambridge University Press 2008).

¹³⁷ Emilian Mihailov and Julian Savulescu, 'Social Policy and Cognitive Enhancement : Lessons From Chess' (2018) 11 *Neuroethics*.

need to ‘effortfully overcome obstacles to peak performance’ and undermines the admirable nature of athletic achievements.¹³⁸

- **Authenticity**: The addition of the statement, ‘the spirit of sport is expressed in how we play true’¹³⁹ implies and reinforces ethical claims based on ‘authenticity’ or ‘no cheating’. Indeed, WADA has, since 2014, celebrated annually a ‘Play True Day’ dedicated to ‘clean sports’.¹⁴⁰ In philosophical terms, the ‘play true’ theory objects to cheating through doping, as the athlete no longer remains their authentic self.¹⁴¹ This objection is associated with the fear of the ‘creation of inauthentic personalities’ and of changing our minds in such a way ‘that it will be difficult for us to attribute moral accountability’ and, as a corollary, praise the cheating athlete.¹⁴² This theory regards success through external aid as less admirable and talent ‘bought’ less worthy, and holds that sometimes it is ‘better to respect “the given” than to try to better things using human abilities’.¹⁴³ Another common denominator of the objection is a ‘general unease about making “the essence of human nature” a project of technological mastery’.¹⁴⁴
- **Equality and Fairness**: The addition of a new paragraph refers to ‘anti-doping programs [that] seek to maintain the integrity of sport in terms of respect for rules, other competitors, fair competition, a level playing field’.¹⁴⁵ This addition appears to reinforce the oft-discussed ethical concept of ‘equality and fairness’. This ethical justification holds that doping in sports erodes the equality and fairness of competition by enabling one competitor to have an advantage over others. WADA

¹³⁸ Erler (n 10) 192.

¹³⁹ WADC (n 45).

¹⁴⁰ ‘WADA Play True Day 2022’ (*WADA*, 2022)

<<https://www.wada-ama.org/en/play-true-day>> accessed 29 April 2022.

¹⁴¹ Michael J Sandel, ‘What’s Wrong with Enhancement’ (2002) *The President’s Council on Bioethics*.

¹⁴² Alberto Carrio Sampedro and José Luis Pérez Triviño, ‘On the Compatibility of Brain Enhancement and the Internal Values of Sport’ (2017) 11 *Sport, Ethics and Philosophy* 316.

¹⁴³ Bostrom and Sandberg (n 7) 327.

¹⁴⁴ *ibid* 326.

¹⁴⁵ WADC (n 45).

is committed to a two-tiered version of fairness: equal opportunity and fair competition. This notion of equality has lately been reflected in WADA's commitment to 'raise the game' through a modernised logo in 2022 in the form of a 'square that represents the rules and customs of the game and the equal sign that evokes fairness with a human touch' (*Figure 3*).¹⁴⁶



Figure 3

Taking these principles at face value, an argument for the prohibition of tDCS by WADA may be constructed as follows. First, given the evidence cited above in relation to the improvement of mental stamina and chess, WADA may argue that the increased time until exhaustion felt by chess players means that they will not need to work as hard to achieve the same results. A similar conclusion can be reached from tDCS studies on cycling¹⁴⁷ and swimming¹⁴⁸ which showed evidence that tDCS can help reduce (the perception of) exhaustion. Therefore, on some level, it appears that tDCS reduces the need for hard work. Secondly, it might be argued that, because tDCS has the potential to enhance, athletes would not 'play true' because they are no longer performing to the best of their innate abilities but rather performing only as well as tDCS allows them to. Finally, it might be argued that the use of tDCS lends itself open to the challenges of enabling an athlete to gain an unfair advantage — a chess-player who uses a tDCS headset to increase mental stamina arguably gains an advantage over other non-users. In this manner, without further scrutinising the merits of the spirit of sport criterion itself, this criterion appears fulfilled. Thus, it is reasonable to speculate that, as tDCS has the *potential* to enhance

¹⁴⁶ 'WADA Unveils Refreshed Brand' (*WADA*, 2022)

<<https://www.wada-ama.org/en/news/wada-unveils-refreshed-brand-launch-its-new-website>> accessed 29 April 2022.

¹⁴⁷ Marcelo Vitor-Costa and others, 'Improving Cycling Performance: Transcranial Direct Current Stimulation Increases Time to Exhaustion in Cycling' (2015) 10(12) PLoS One; Alexandre Hideki Okano and others, 'Brain Stimulation Modulates The Autonomic Nervous System, Rating Of Perceived Exertion And Performance During Maximal Exercise' (2013) 49 British Journal of Sports Medicine.

¹⁴⁸ Salehi and others (n 133).

performance, and as it seemingly flouts the spirit of sport as defined by WADA, it is likely that tDCS and similar technologies would be prohibited based on the two-thirds criteria outlined above.

However, in the following Section, this article will argue that the above reasoning is untenable and that these underlying ethical principles are unconvincing at best and deeply flawed at worst — *even* in the context of physical sports in which they were originally articulated. For now, it is sufficient to say the following: the fact that under WADA's current regulatory framework tDCS can be taken to satisfy the two-third criteria for being added to PLIS (by virtue of the fact that it can potentially fulfil both the enhancement and the spirit of sport criteria) suggests deeply rooted issues within WADA's framework.

II. DECONSTRUCTING WADA'S REGULATIONS

Section I highlighted four key problematic areas in WADA's anti-doping regulations and concluded by establishing that tDCS could, in theory, come within WADA's purview. This Section will now seek to disentangle these issues and discuss the problems that tDCS and chess help bring to the fore.

2.1 The Mental/Physical Divide

As highlighted in Section 1.3, the compilation of PLIS in accordance with Article 4.3 of WADC makes no distinction between mental and physical sports, and the illustrative example of tDCS in chess accentuates WADA's deficiencies in the following manner.

First, if WADA indeed imposes a ban on tDCS, all forms of sport would be caught as no distinction is drawn between physical and mental sports. This is problematic because *different sports have different values*. By imposing a blanket ban, WADA would fail to recognise the core values of (and hence the spirit of sport unique to) that particular sport which WADA purports to protect in the first place.

Although the ostensible purpose of every sport is to compete and win, the values underlying each sport (and therefore the corresponding ethical

justifications for regulating doping for that sport) are different. For example, whilst the core values of marathon running are resilience, determination and stamina, in other sports, such as gymnastics, there is an emphasis on poise, flexibility and agility; in shooting, what is valued is accuracy, steadiness and consistency. Values of chess are arguably more difficult to pinpoint, with former world chess champion Anatoly Karpov once remarking that '[c]hess is everything: art, science and sport'.¹⁴⁹ Notwithstanding, it appears that 'the value of chess lies primarily with the intellectual and creative'¹⁵⁰ pursuit of excellence and, as such, showcasing superior (physical) stamina is not the primary purpose of chess. Shaw provides a useful comparison: 'the value of running a marathon is more closely connected with stamina and a marathon runner who took a physical stamina enhancement might be accused of missing/undermining the point of marathons ... The value of a work of art, such as a painting, [on the other hand] does not seem to depend on how the artist got the energy to create it'.¹⁵¹ In response to the potential criticism that chess is not directly comparative to art, Shaw further uses performing art, such as ballet, for comparison.¹⁵² Whilst part of the appreciation of ballet is the stamina exhibited by ballerinas (in being able to dance for long periods of time), stamina is not the primary reason for appreciation. Indeed, the same can be said about chess — the appreciation of chess does not lie in *how* the chess-player obtains their energy to last through the tournament.

At this point, an important distinction must be made between two overlapping but crucially distinct concepts: 'skills' and 'values'. The former refers to the ability to perform an action, typically learned or acquired through practice or training; for example, balancing, skipping or throwing in physical sports. In mind sports, this can be arithmetic ability, hand-eye coordination and even mental stamina. The concept of 'values', on the other hand, denotes the abilities and qualities necessary for the successful pursuit of a particular sport (e.g., resilience, teamwork and commitment). It is argued here that any improvement of mental stamina via the use of tDCS which may have the potential to enhance chess performance (see Section 1B) should not be confused with or taken to undermine the central values of chess, including its

¹⁴⁹ Krzysztof Olechnicki, 'The sociological paradox of chess: the transclass distinction of the game of kings' (2020) *Acta Universitatis Lodziensis Folia Sociologica* 49, 58.

¹⁵⁰ Shaw (n 80) 222.

¹⁵¹ *ibid.*

¹⁵² *ibid.*

value as an intellectual and creative pursuit requiring determination, perseverance, and hard work — as discussed in detail in Section 2.3(i).

Going one step further, it has been argued by some that cognitive enhancement cannot make a professional chess-player much better at the game.¹⁵³ Golf has noted that cognitive enhancement drugs could only have effects ‘with persons at low cognitive baseline’¹⁵⁴ and that, with ‘normal’ persons, cognitive enhancement is ‘still below clinical significance’. Hence, it is at least doubtful that cognitive enhancement drugs as currently constituted could enhance chess expertise, much less so at a professional level.¹⁵⁵ Instead, Golf has suggested that 60% of the FIDE rating of a chess-player is attributable to these variables: the player’s age of entry into a chess club, the number of tournament games played, emotion expression control, numerical intelligence, actual age and, finally, chess-related performance motivation.¹⁵⁶ To add to this list, an effective nutrition regime can help accumulate ‘super compensated glycogen’ in the brain, muscle and liver.¹⁵⁷ Perhaps it is these factors which have led to many professional chess players feeling insulted at the insinuation that doping in chess is possible. Indeed, since 2002, top chess players have been protesting against the introduction of doping tests in chess by FIDE, with GM Artur Jussupow, once German champion, declaring that FIDE’s lack of clear definitions meant that such tests placed players “at the mercy of FIDE” and “remind[ed] [him] of the Stalin methods during the thirties”.¹⁵⁸ Having conducted experiments ‘for around twenty years’, Dr Helmut Pfleger, himself both a medical doctor and a chess grandmaster, also expressed doubts about the correlation between doping and chess, stating that “both mentally stimulating and mentally calming medication have too many negative side effects”.¹⁵⁹ In sum, it is argued that doping cannot give a pronounced advantage to professional chess players and cannot replace hard work as the pathway to success.

¹⁵³ Mihailov and Savulescu (n 138).

¹⁵⁴ Sighart Golf, ‘Doping For Chess Performance’ (2015) 5(3) *Journal of Sports Medicine & Doping Studies*.

¹⁵⁵ Golf (n 111) 5.

¹⁵⁶ Golf (n 155).

¹⁵⁷ *ibid*.

¹⁵⁸ ‘Controversy Over FIDE Doping Check’ (*Chessbase*, 2002)

<<https://en.chessbase.com/post/controversy-over-fide-doping-check>> accessed 27 September 2022.

¹⁵⁹ *ibid*.

Turning now to WADA's lack of discernment regarding the physical/mental divide and the idiosyncrasies of each sport, the risk of overinclusion is significant. The problem with such a lack of sensitivity is that a ban on tDCS and other cognitive enhancers would apply indiscriminately to all sports, regardless of the differing core values, even if the core values of a sport are unlikely to be challenged by the use of tDCS. As argued above, in the case of chess, improvement of mental stamina would not undermine the values of what makes a good chess-player. Since the 'ability to keep playing for a long time without becoming fatigued does not seem central to the *value* of chess',¹⁶⁰ the value of chess as a pursuit of excellence is not fundamentally challenged by the use of tDCS or any associated enhancement of mental stamina, and thus the central purpose and values of chess are preserved.¹⁶¹ Doping regulations and policies formulated without regard to the true values of a sport have the benefit of easy administration but are founded on rather indefensible and precarious grounds.

Furthermore, not only does WADA's over-inclusive regulation fail to capture the differences between physical and mind sports, but it is also poorly, if at all, suited to capture the nuances between the different types of mind sports themselves. Esports is a good example. Whilst the objective in chess is relatively straightforward — checkmating your opponent — the objectives in different Esports range from demonstrating superior reaction times, memory recall, or concentration power. This effectively means that, as *between* mind sports (that is, chess versus Esports) and even *within* a particular mind sport (that is, Esports), different sets of core values dominate, making indiscriminate application of anti-doping rules all the more problematic. Relatedly, over-inclusion, or a blanket prohibition, by WADA could be criticised as artificial — a source of mistrust and fear. Mihailov and Savulescu are correct in pointing out that regulations based on the lack of the physical/cognitive divide are at risk of 'creating a detrimental culture of suspicion that ascribes unwarranted blame'.¹⁶²

¹⁶⁰ Shaw (n 80) 222.

¹⁶¹ *ibid* 220.

¹⁶² Mihailov and Savulescu (n 138) 116.

2.2 Cognitive Enhancement

Hitherto, the regulation of WADA's cognitive enhancement through PLIS compiled pursuant to Article 4.3 has very much been an afterthought as the construction of the framework has largely been responsive to scandals around physical sports. Furthermore, because of WADA's deontological framework, there has been, conveniently, no reason to examine the nuances involved in cognitive enhancements, including *how* and *to what degree* a cognitive enhancement may impact either a physical or a mind sport. The focus on whether there is *potential* to enhance means that WADA overlooks the fact that, unlike physical enhancement, cognitive enhancement generally does not operate in a linear fashion.

This is particularly pertinent to chess and tDCS because chess performance cannot be tied to monolithic factors and is instead influenced by a multitude of complex variables. For example, whilst the above has established that tDCS may help with the mental stamina of a chess-player, this alone is not enough to guarantee improved performance. As mentioned above, other relevant factors include attention and information processing. Mind sports, as opposed to physical sports, are characterised (and powered) by cognitive complexity composed of a variety of specialised cognitive abilities. Relative to physical sports, the correlation between enhancement and intervention in mind sports is more difficult to establish. For example, empirical evidence of increased oxygen supply post-blood doping leading to an increase in the power of endurance can be more easily established than the correlation between the quality of a chess move pre and post-tDCS usage, bearing in mind that the particular move (or, effectively, the combination of moves as chess players calculate prior to taking the *first* of a *series* of moves) is the sum of processing, calculation, and memory recall in addition to the ability to withstand fatigue. Therefore, in the example of tDCS and chess, the current deontological approach of WADA under Art 4.3.1 demands that tDCS be banned because it has the *potential* to enhance mental stamina — but this is not the same as saying that tDCS *can* enhance chess performance. Mental stamina is only one of a number of pertinent factors.

Chess and tDCS further expose the deontological approach in this way: it is conceivable that enhancement of one cognitive aspect, for example, alertness (brought on by an increase in mental stamina via tDCS), can be *detrimental* to the overall performance as it impacts on other factors, such as the chess-player's faculty of information processing, since increased alertness may cause players to become hyper-focused on one point resulting in a net loss of performance. Indeed, a 2017 study¹⁶³ showed that modafinil has the ability to calm the nerves of chess players *but* ultimately caused them to lose on time due to slower play and prolonged thinking brought on by the drug. In this manner, a deontological approach would ban tDCS *despite* the fact that it *could* be detrimental to the overall performance.

In conclusion, although cognitive enhancement brought on by NIBS techniques can indeed enhance certain aspects of performance, the complexity of mind sports means that such techniques cannot enhance holistically. Chess and tDCS have demonstrated that mental stamina enhancement is only one facet of a larger cognitive process and, importantly, such an enhancement would not violate the core value of chess as seen in Section 2.1, making a blanket ban unreasonable and patronising. WADA must therefore acknowledge the nuances of cognitive enhancement: a *consequential* approach based on *actual* enhancement is more suitable and just — a point to which this article will return in the final Section.

This article now turns to the fundamental ethical justifications for prohibiting doping.

2.3 Spirit of Sport

In Section 1.5(c), the argument that the spirit of sport criterion might be violated by using tDCS in the context of chess was discussed, relying on WADA's principles of hard work, authenticity, and fairness. At the same time, the strength of that argument was questioned, which this article will now seek to analyse. This Section will aim to deconstruct these principles and examine how the policies underpinning anti-doping rules fail not only to live up to expectations but to stand on their own merits. Furthermore, the use of tDCS in

¹⁶³ Franke and others (n 37).

the context of chess prompts additional and pertinent philosophical considerations that may further undermine the validity of these principles as constitutive of the very foundation of the spirit of sport. However, it should be noted that, since these principles underpinning the anti-doping regulations themselves make no distinction between physical and mental sports, it would not be appropriate to confine this discussion to a pure mind sport/chess context; instead, broader references, where relevant, will be made to physical sports to achieve a holistic discussion.

i) Enhancement does not necessarily reduce the need for hard work

The assumption that tDCS *removes* the need for hard work, and therefore is an unacceptable intervention, is unsubstantiated. In fact, the opposite is true — tDCS may actually *enhance* hard work.

In the context of chess, the ergogenic effects of tDCS cannot take away the value of effort which enables the player to attain his/her FIDE rating in the first place. Charness et al. confirm the correlation of ‘time spent on serious study alone, tournament play, and formal instruction’ with ‘chess skills as measured by tournament performance ratings’.¹⁶⁴ Significantly, it is the amount of training time, *not* intelligence, that is ‘one of the most effective prognostic factor[s] for achieving chess expertise’.¹⁶⁵ Importantly, tDCS does not reduce the amount of training time needed to reach the same performance level by enhancing mental capabilities. Instead, tDCS enables chess players to train for longer — a subtle yet important distinction. If tDCS were to be able to cut short the time needed for training, this would be equivalent to a ‘shortcut’. However, it is suggested here that this is not the case — the time invested to reach different levels of expertise remains unchanged. To reach their level of competence, it has been estimated that grandmasters spend at least 5,000 hours of chess training time ‘during the[ir] first decade of serious chess play’,¹⁶⁶ a critical period where studies have shown that ‘expos[ure] to a chess environment at an early age ... is important for development skills’.¹⁶⁷ Specifically, it has been

¹⁶⁴ Neil Charness and others, ‘The Role of Deliberate Practice in Chess Expertise’ (2005) 19(2) *Applied Cognitive Psychology* 151.

¹⁶⁵ Golf (n 111).

¹⁶⁶ Charness and others (n 65).

¹⁶⁷ Fernand Gobet and Guillermo Campitelli, ‘The Role of Domain-Specific Practice, Handedness, and Starting Age in Chess.’ (2007) 43 *Developmental Psychology* 27.

estimated that 'unrated players reported more than 8,000 hours; rated but untitled players reported close to 12,000 hours; FIDE Masters almost 20,000 hours and International Masters 28,000 hours of dedication to chess'.¹⁶⁸ These numbers paint a clear picture: expertise in chess cannot be detached from the hours spent practising. Even with the use of tDCS, the same number of hours would need to be invested. In terms of pattern recognition alone, it has been commented ¹⁶⁹ that 'a very good chess player remembers at least 300,000 chess positions' prior to being 'able to recognize and then to compare and to act'.¹⁷⁰ Computer program MAPP calculated the need to acquire up to 100,000 patterns¹⁷¹ to reach Master level, while another computer program CHREST shows that 'at least 300,000 chunks are required to reach grandmaster level'.¹⁷² Mathematically speaking, tDCS can in no way substitute the effort required to acquire hundreds of thousands of patterns within tens of thousands of hours of practice — there is simply no shortcut to the hard work required and there can certainly be no 'quick fix'.¹⁷³ tDCS cannot replace the amount of time and effort that goes into becoming an elite chess-player. Effort, in many ways, is simply unavoidable.

In fact, building on this argument, an unqualified concept of 'hard work' as a basis for the anti-doping sentiment is problematic — there is an argument to be made that the use of tDCS is in fact *evidence* of hard work. tDCS could allow a chess-player to devote more time to 'tackling challenges that are inherent in worthwhile activities',¹⁷⁴ such as searching for the perfect move when in competition or solving chess puzzles, learning new openings or improving endgame tactics when in training. Indeed, Shaw has gone as far as to argue that tDCS can 'enable professional chess-players to practise for longer, thereby leading to more effort and perseverance overall'.¹⁷⁵ In agreement,

¹⁶⁸ Golf (n 111) 5.

¹⁶⁹ *ibid* 3.

¹⁷⁰ *ibid* 3.

¹⁷¹ Herbert A Simon and Kevin Gilmarin, 'A simulation of memory for chess positions' (1973) 5 *Cognitive Psychology* 29-46.

¹⁷² Fernand Gobet and Neil Charness 'Expertise in Chess' [2006] *The Cambridge Handbook of Expertise and Expert Performance* 11.

¹⁷³ Shaw (n 80) 221.

¹⁷⁴ *ibid* 222.

¹⁷⁵ *ibid* 221.

Maslen et al. have further added¹⁷⁶ that the choice to use a less arduous route in achieving the same goal may in fact demonstrate greater commitment and hard work. In this manner, tDCS does not detract from the hard work required but should be considered a reduction in principle of the perception of the ‘influence of mental barriers to top performance’.¹⁷⁷ By pushing harder, an athlete’s chances of maximising performance are increased. A counterargument may be leveled here that tDCS allows chess players to achieve the same level quicker than they would have been able to without the technique because they can train for longer. Two responses here are offered: first, there is still actual effort, and it is important to distinguish this from perceived effort, or indeed no effort. By analogy, it has been noted that the use of timetables in studying may decrease the overall effort needed and make the task more efficient by ‘externally control[ling] one’s weakness of will’.¹⁷⁸ Likewise, where a chess-player chooses to use more efficient methods of training, like tDCS, this is not fundamentally different from setting a timetable for studying with the goal of increasing productivity,¹⁷⁹ which in truth reflects the player’s commitment and dedication to perfecting their game. Such time can be used, for example, to engage with cognitive training aids such as ‘chess boards, chess books, building chess images, visual observation of chess games, vocational training with chess, metacognitive training’.¹⁸⁰ Secondly, if chess players can achieve the same level quicker, this allows them to exert more time to improve their chess game *further*, elevating the overall standard in chess. In turn, to use WADA’s own terms, tDCS can in fact facilitate the ‘pursuit of human excellence’.¹⁸¹

A final observation must be made here in relation to how we understand social acceptability of drugs in performance. The ergogenic effects of tDCS on chess players can be likened to those of beta-blockers used by classical musicians. Beta-blockers can lower heart rate, blood pressure, and control stage fright. In the context of music, it can help classical musicians express their skills better, resulting in a deserving performance reflective of their

¹⁷⁶ Hannah Maslen, Julian Savulescu and Carin Hunt, ‘Praiseworthiness and Motivational Enhancement: ‘No Pain, No Praise?’ (2019) 98 *Australasian Journal of Philosophy* 304-318.

¹⁷⁷ Erler (n 10).

¹⁷⁸ Maslen and others (n 176).

¹⁷⁹ Shaw (n 80).

¹⁸⁰ Golf (n 155).

¹⁸¹ WADC (n 45).

true ability.¹⁸² Taking this one step further, arguably, the most famous band ever, the Beatles, wrote entire albums inspired by their experimentation with narcotics and psychedelics.¹⁸³ In the context of visual arts, the use of narcotics such as opium was at one point common, with Pablo Picasso perhaps being the most prominent example.¹⁸⁴ The point here is that, on some level, the use of narcotics and other substances (which may or may not offer an immediate advantage and may even at times be detrimental to performance) as an *enhancement* of performance has historically been a socially acceptable, if not encouraged, phenomenon. The evident riposte to this argument is that arts are unlike sports, for the latter focuses on the competitive aspect, allowing only one winner. Three responses can be offered. Firstly, whilst there is no ‘winner’ or ‘loser’ in relation to arts, it is nonetheless often competitive with equally large stakes at play. Secondly, battling fatigue is not the central focus of the spirit of sport in chess, just as battling stage fright is not the essence of musical performances. Thirdly, and more fundamentally, on a closer reflection of WADA’s own terms of the ‘dedicated perfection of each Athlete’s natural talents’,¹⁸⁵ there is an emphasis on the amplification of talent that an athlete *already* possesses. Much like how narcotics and psychedelics may have inspired the Beatles to write the hit song ‘Lucy in the Sky with Diamonds’ (in its abbreviated form — LSD — also the abbreviation for lysergic acid diethylamide, a psychedelic drug), and how opium inspired Picasso’s paintings, such as the ‘Family of Saltimbanques’,¹⁸⁶ the use of tDCS might be said to amplify the talent *already* in existence. Therefore, rather than removing the need for hard work, tDCS, just like the use of narcotics, has the potential to allow the true expression of the ‘natural talent’ of a chess-player. This argument of ‘true expression’ will be further explored below in the discussion on ‘authenticity’, to which we now turn.

¹⁸² Savulescu and others (n 89) 667.

¹⁸³ Mikal Gilmore, ‘Beatles’ Acid Test: How LSD Opened the Door to ‘Revolver’ (Rolling Stone, 25 August 2016)

<<https://www.rollingstone.com/feature/beatles-acid-test-how-ld-opened-the-door-to-revolver-251417/>> accessed 29 April 2022.

¹⁸⁴ Jeff Goldberg, ‘The Complicated Relationship Between Opium and Art in the 20th Century’ (Arty, 4 May 2018)

<<https://www.artsy.net/article/artsy-editorial-complicated-relationship-opium-art-20th-century>> accessed 29 April 2022.

¹⁸⁵ WADC (n 45).

¹⁸⁶ Goldberg (n 185).

ii) Enhancement does not necessitate a loss of authenticity

The claim that tDCS and other enhancements might create 'inauthentic personalities' falls into the fallacy of assuming that there exists an 'authentic self' which would be lost if external aid were to be relied upon. Yet, clearly, we would not consider an individual with a prosthetic leg to have lost moral agency, nor would we call someone in a wheelchair a person who has lost culpability or human freedom. Their equipment/treatment (i.e., 'enhancement') makes them no less 'human'. The likely response is that a wheelchair or a prosthetic is a therapy, not an enhancement, the idea being that interventions should be allowed for 'treatments responding to genuine medical needs' but not for enhancements that 'satisfy mere preferences or expensive tastes'.¹⁸⁷ Indeed, WADA recognizes the therapeutic value of banned substances or methods and outlines rules for therapeutic use exemptions (TUE), which essentially allows an athlete to use a medication or method if certain criteria are met including its use for treatment of a diagnosed medical condition and there exists 'no reasonable, permitted, alternative treatments'.¹⁸⁸ For instance, the famous gymnast Simone Biles utilised TUE in Rio 2016 for Ritalin (methylphenidate), a banned substance,¹⁸⁹ which she took to manage her ADHD symptoms.¹⁹⁰

This therapy/enhancement line serves a Herculean function: that of demarcating the legitimate from the illegitimate. It is tasked with assisting the important ethical adjudication of *when* individual choice and autonomy to engage with a technology should be allowed to prevail over regulations.¹⁹¹ The key question in this context is whether a line can be drawn in any principled fashion. This article suggests that it cannot.

¹⁸⁷ Alexandre Erler, 'The limits of the treatment-enhancement distinction as a guide to public policy' (2017) 31 *Bioethics*.

¹⁸⁸ WADA TUE (n 128).

¹⁸⁹ PILS (n 108).

¹⁹⁰ Meredith Cash, 'Simone Biles' ADHD medication is banned in Japan, but a Tokyo 2020 exemption allowed them for athletes' (*Insider*, 30 July 2021), <<https://www.insider.com/simone-biles-adhd-meds-banned-japan-impacting-performance-2021-7>> accessed 26 November 2022.

¹⁹¹ Isabel Karpin and Roxanne Mykitiuk 'Going out on a Limb: Prosthetics, Normalcy and Disputing the Therapy/Enhancement Distinction' (2008) *Medical Law Review* 16.3, 413-436.

The difficulty with drawing such a distinction is especially highlighted by the introduction of NIBS technologies such as tDCS, given the growing recognition of therapeutic use of tDCS in areas such as MDD, with the National Institute for Health and Care Excellence endorsing it as a treatment for depression.¹⁹² It is conceivable that if, in time, tDCS were to be banned, an elite chess-player with MDD could apply through FIDE's TUE scheme¹⁹³ and obtain the right to use it with all the ergogenic effects tDCS potentially brings. In such circumstances, the use of tDCS is restorative — depression should not inhibit one's ability to perform in chess tournaments. On the other hand, if we accept the premise that tDCS can enhance performance in chess, would one say that a depressed chess-player using tDCS has gained an advantage that they would not otherwise have had and therefore be regarded as 'inauthentic'? This example illustrates the difficulty in drawing a neat distinction between enhancement and treatment.

This observation in the context of tDCS can also be made in the context of various other sports as well. Returning to gymnast Biles, would one say, for example, that the improved cognitive power of Biles to concentrate balances with the goal of restoring her therapeutically? Alternatively, one might take a view that it is possible Ritalin aided Biles in her successes by enhancing her skills to achieve a score she would not otherwise have obtained. In addition, one must consider where this line should be drawn: should TUE be allowed to restore Biles to the level of an average gymnast (with normal concentration levels), or that of an Olympic gymnast (presumably, with superior concentration power)? Arguably, if we understand WADA's notion of 'natural talents', the latter must be true — Biles should be able to express her true talent without being hindered by ADHD, which is beyond her control. On the other hand, some would contend that ADHD is just like any other defect, such as height, weight, foot size, propensity for nerves, etc, that may affect a gymnast's performance. In this sense, Biles' ADHD should be considered as part of her 'authentic' self, and only by competing with naturally given talent can an athlete 'play true'. This

¹⁹² NICE, 'Transcranial direct current stimulation (tDCS) for depression' (NICE, 16 August 2015), <https://www.nice.org.uk/guidance/ipp530/chapter/1-recommendations> accessed 29 April 2022, pt1.

¹⁹³ FIDE, 'Therapeutic Use Exemptions application form' (FIDE Handbook, 2018) https://handbook.fide.com/files/handbook/TUE_application_form.pdf accessed 29 April 2022.

tension in where the line should be drawn is a fallacy because of the assumption that such a line does actually exist. There can be no such line and, indeed, some have argued that an assertion of its existence is merely ‘shifting lines in the sand’.¹⁹⁴ Indeed this ‘line’ between treatment and enhancement is an impractical one to draw, and therefore an inadvisable one to use.

Take another example of Oscar Pistorius, the champion paralympic sprinter fitted with blades who applied to run in the able-bodied Olympics and succeeded. Were his blades a treatment or an enhancement, or were they once a treatment turned into an enhancement? It is impossible to accurately answer this question as to whether an advantage above what is therapeutically ‘acceptable’ was gained. For this reason, the decision to allow Pistorius to run in the 2012 London Olympics was not without controversy,¹⁹⁵ especially given that the issue of whether the prosthesis ‘provided him with an overall net advantage or disadvantage’¹⁹⁶ was never explored at the hearing. The permitted use of a pair of J-shaped ‘Cheetah’ blades contrasts sharply with a small pill innocently taken for therapeutic purposes to treat common cold by 16-year-old Romanian gymnast Andreea Raducan at the order of her team doctor.¹⁹⁷ Unlike the fortune of Pistorius, Raducan was stripped of the coveted all-round Women’s Individual Olympic title.¹⁹⁸ In the words of the then Romanian Olympic Committee President Ion Tiriatic, the medicine was “not enhancing but diminishing [to] performance”.¹⁹⁹ The differing treatment of these athletes raises the question of

¹⁹⁴ Patrick Lin, ‘Therapy and Enhancement: Is There a Moral Difference?’ (*Genetic Engineering and Biotechnology News*, 1 July 2009)

<<https://www.genengnews.com/magazine/116/therapy-and-enhancement-is-there-a-moral-difference/>> accessed 29 April 2022.

¹⁹⁵ Jeré Longman, ‘An Amputee Sprinter: Is He Disabled or Too-Abled?’ (*The New York Times*, 15 May 2007)

<<https://www.nytimes.com/2007/05/15/sports/othersports/15runner.html>> accessed 30 April 2022.

¹⁹⁶ Court of Arbitration for Sport *Pistorius v IAAF* (2008) [37].

¹⁹⁷ Nick Zaccardi, ‘Andreea Raducan will not be reinstated as 2000 Olympic all-around champion after meeting with IOC president’ <<https://olympics.nbcsports.com/2015/06/25/andreea-raducan-olympics-medal-stripped-gymnastics-sydney-2000-romania/>> accessed 7 October 2022.

¹⁹⁸ *ibid.*

¹⁹⁹ Amy Shipley, ‘Romanian Gymnast Stripped of Gold’ (*The Washington Post*, 26 September 2000)

<<https://www.washingtonpost.com/archive/sports/2000/09/26/romanian-gymnast-stripped-of-gold/e8c4f17a-a86a-41ce-b281-9d71b496d96f/>> accessed 1 May 2022.

where this line between enhancement and treatment should be drawn, and whether the drawing of such a line is ever even possible.

The harsh result from the Raducan case is also further evidence of the absurd decisions reached by WADA's strict regime. Their focus on the behaviour of athletes rather than the outcome is in fact harmful to authenticity. Whilst it is beyond the scope of this essay to delve into the academic intricacies of the 'thick' and 'thin' conceptions of authenticity, one can view authenticity for what it is not: where one's true character (i.e., the qualities unique to an individual) is being suppressed, they cannot be acting 'authentically'. In Raducan's case, it was clear that the common cold had indeed suppressed her ability to perform at her best, and to punish her for attempting to perform authentically was a hypocritical and retrograde stance. Even if the pill Raducan had taken did enhance her performance, this is not substantively different from Biles' use of Ritalin. The only clear distinction is that Biles had applied for TUE and Raducan had not. As such, what WADA is punishing the athlete for is not the enhancement that affects authenticity or some substantive unfairness created but administrative errors. It is then arguable that this is not the behaviour to be expected from a body purporting to protect the rights of athletes and the fairness of competition.

At this juncture, of relevance is WADA's interpretation of 'natural talents' underlying the concept of authenticity, which is questionable when considered in the context of the growing transgender/hyperandrogenic classification debate. In this debate, there are, on one hand, hyperandrogenic athletes, such as Caster Semenya, who possess higher than normal levels of testosterone and, on the other, transgender athletes, such as Lia Thomas. In relation to Semenya, even though her gender is 'uncontroversially female' and that she is legally so regarded,²⁰⁰ she has consistently faced hurdles, notably, the World Athletics announcing in 2018 rules that prohibit athletes with disorders of sex development (DSD) from competing.²⁰¹ In April 2019, CAS dismissed

²⁰⁰ Julian Savulescu, 'Ten Ethical Flaws In The Caster Semenya Decision On Intersex In Sport' (*The Conversation*, 2019) <<https://theconversation.com/ten-ethical-flaws-in-the-caster-semenya-decision-on-intersex-in-sport-116448>> accessed 26 September 2022.

²⁰¹ Court of Arbitration for Sport, 'Executive Summary Caster Semenya, Athletics South Africa, IAAF' (*CAS*, 2019)

Semenya's request for arbitration on the basis that she had failed to establish that the then IAAF's eligibility classifications were 'invalid'²⁰² (CAS' Judgment) and further that, although the regulations were found 'discriminatory', such discrimination was a 'necessary, reasonable and proportionate means of achieving the legitimate objective of ensuring fair competition in female athletics'.²⁰³ Although this was CAS' judgment and not WADA's, notably, WADA did commission a study²⁰⁴ which suggested that hyperandrogenic athletes do have a competitive advantage. In this manner, it appears that WADA likely agrees with the IAAF's conclusion that Semenya should take testosterone suppressing substances.

Here, it appears that WADA does not seem to care as much about what comes naturally as it does about what is being *enhanced*, as evidenced by the stipulation to suppress what seems to be the enhancer — testosterone. In reality, other forms of genetic predispositions that enhance athletes' performance are not uncommon and there is evidence to support the claim that record-breaking athletes are often genetic outliers. This idea is explored more in the next subsection below, but some simple examples here will help illustrate the point: Lance Armstrong had a 'preternaturally high maximum oxygen consumption and superbly efficient oxygen use',²⁰⁵ volleyball great Flo Hyman had Marfan syndrome 'that gave her the tall stature and long arms that likely contributed to her success'²⁰⁶ and, at the time of writing, only four of the top 50 marathon times in the world are not achieved by runners from Kenya or Ethiopia²⁰⁷ — an impressive feat attributable to their 'slight, lithe physiques and exposure to high

<https://www.tas-cas.org/fileadmin/user_upload/CAS_Executive_Summary__5794_.pdf> accessed 28 September 2022. ('CAS Executive Summary').

²⁰² *ibid.*

²⁰³ *ibid.* 1.

²⁰⁴ 'IAAF To Use Study Saying Female Athletes Get Large Testosterone Boost In Rule Appeal - Olympictalk | NBC Sports' (*OlympicTalk | NBC Sports*, 2017)

<<https://olympics.nbcsports.com/2017/07/03/female-track-and-field-athletes-testosterone/>> accessed 26 September 2022.

²⁰⁵ J. Michael Bostwick and Michael J. Joyner, 'The Limits Of Acceptable Biological Variation In Elite Athletes: Should Sex Ambiguity Be Treated Differently From Other Advantageous Genetic Traits?' (2012) 87 Mayo Clinic Proceedings.

²⁰⁶ *ibid.*

²⁰⁷ 'World Athletics | Marathon - Men - Senior - Outdoor' (*worldathletics.org*, 2022)

<<https://www.worldathletics.org/records/all-time-toplists/road-running/marathon/outdoor/men/senior>> accessed 27 September 2022.

altitudes'.²⁰⁸ If WADA does not take issue with such natural biological endowments/enhancements as being prohibitive to the 'authentic' self, then surely Semenya should equally be free to express herself authentically. The stance WADA and the World Athletics have taken — that higher levels of testosterone are detrimental to authentic performance — is self-conflicting given that similar prohibitions do not exist for athletes with other forms of 'superior' biological traits. A full discussion on the legal and ethical implications of CAS' judgment to uphold the World Athletics' ban on Semenya²⁰⁹ is beyond the scope of this article, but suffice it to say that WADA's implicit approval of the World Athletics' stance offers an inconsistent and untenable view on how they interpret 'natural talent'.

Transgender athletes, such as Lia Thomas, can be distinguished from athletes such as Semenya. Whilst Semenya was born with a naturally higher testosterone level relative to other female athletes, elevated levels of testosterone in Thomas were a result of her conscious decision to transition. It has been argued that this deliberate therapy-turned-enhancement gave her an unfair competitive advantage and led to her 500m freestyle victory at the NCAA's Women Championship (although this assertion that she actually gained an advantage has been challenged²¹⁰). In this situation, regardless of whether her performance was actually enhanced, it can be argued that the transition had allowed her to be her *authentic* self. Thomas herself recently commented that, "I transitioned to be happy, to be true to myself".²¹¹ In this scenario, what is true is that *enhancement enabled authenticity*. Whether or not it is *fair* for Thomas or other transgender athletes to swim in the female category is beyond the scope of this discussion, particularly since Thomas and the NCAA fall outside WADA's purview. The pertinent point to stress, however, is this: in the case of Semenya, her 'enhancement' is her authentic self and, in the case of Thomas, her 'enhancement' *allowed* her to reach her authentic self. In both cases, there are

²⁰⁸ Bostwick and Joyner (n 206).

²⁰⁹ For example, see Savulescu (n 201).

²¹⁰ Io Dodds, 'Critics Accuse Trans Swimming Star Lia Thomas Of Having An Unfair Advantage. The Data Tells A Different Story' (*Independent*, 2022) <<https://www.independent.co.uk/news/world/americas/lia-thomas-trans-swimmer-ron-desantis-b2091218.html>> accessed 26 September 2022.

²¹¹ 'Lia Thomas Says She Transitioned To Be Happy, Not To Win Swimming Titles' (*The Guardian*, 2022) <<https://www.theguardian.com/sport/2022/may/31/lia-thomas-trans-swimmer-university-of-pennsylvania-ncaa-title>> accessed 26 September 2022.

respectable grounds to argue they competed *authentically* within the meaning of that word and thus any sentiment that enhancement is detrimental to authenticity must be flawed.

Paradoxically, outside this therapy/enhancement distinction, it may even be asserted that tDCS can in fact *enhance* the authenticity of athletes and of chess players alike. It is suggested that part of authenticity is ‘self-creation’, the idea that, by attempting to get better via whatever means, is an act of authenticity. This view takes the stance that part of authenticity is aligning oneself with one’s authentic urges. In terms of chess, with a reduction in mental fatigue after the use of tDCS, players in competition are arguably motivated to play even better, to push beyond their limits and to go the extra mile to achieve their ‘authentic selves’. In general, external aids should not be seen as categorically objectionable in the name of authenticity. Furthermore, is it not true that, in many cases, external aids, otherwise known as shortcuts, are already tolerated? For example, spike shoes — which enable a better grip of the running surface — are allowed in sprinting; vibration dampening tennis rackets ‘designed to mitigate frame vibration’²¹² which delay muscle fatigue are allowed in tennis; and maltodextrin, a form of energy gel that enables athletes to ‘maintain ... energy levels until the end of the game’²¹³ is widely used in football. In some cases, external aids may be deemed to have gone too far, for example, Nike’s Alphafly shoes which have been banned by the World Athletics. However, this is an extreme case where the shoes have been estimated to have ‘lowered the energetic cost of running by 4% on average’ and translate to about a ‘3.4% improvement in running velocity’.²¹⁴ Improvements by such amounts have been

²¹² Yeh I-L. and others, ‘Vibration-Damping Technology in Tennis Racquets: Effects on Vibration Transfer to the Arm, Muscle Fatigue and Tennis Performance’ (2019) 1 Sports Medicine and Health Science 49.

²¹³ Dan Sheldon, ‘Southampton Medical Checks, Gels and Team Talks Between 60th -70th Minute: Coincidence Or Plan?’ (*The Athletic*, 2022) <https://theathletic.com/3128906/2022/02/16/southampton-medical-checks-gels-and-team-talks-between-60th-70th-minute-coincidence-or-plan/?source=emp_shared_article> accessed 15 March 2022.

²¹⁴ James Witts, ‘Technological Doping: The Science Of Why Nike Alphaflys Were Banned From The Tokyo Olympics | BBC Science Focus Magazine’ (*Sciencefocus.com*, 2021) <<https://www.sciencefocus.com/the-human-body/nike-alphafly-banned-technological-doping/>> accessed 15 March 2022.

equated to 'technological doping'.²¹⁵ In general, however, external aids have largely been accepted as part of sports.

In the cognitive sphere, laptops are used and specially designed with AI programs facilitating students' learning; in some campuses, students are not forbidden from taking Ritalin for better focus in exam preparations (in addition to other enhancing substances, such as caffeine, snack bars and glucose drinks). In the chess world, chess engines, such as Fritz and ChessBase, are not prohibited and are known to be good training aids for chess players. These, and countless others, are examples of how cognitive enhancements extend and complete a person's talents, promoting authenticity by 'offloading irrelevant, repetitive, or boring tasks' through external aids which 'enabl[e] a person to concentrate on more complex challenges that relate in more interesting ways to his or her goals and interests'.²¹⁶ Bostrom is right when he asserts that, 'insofar as cognitive enhancements amplify the capacities required for autonomous agency and independent judgment, they can help a person lead a more authentic life'.²¹⁷ In a world of rapid technological advances, we should embrace the inevitable changes and the benefits they provide. This does not mean we should indiscriminately accept them; rather, parameters need to be drawn for their appropriate application. Accordingly, the time is ripe for WADA to do the same for tDCS.

In this manner, it is difficult to be convinced that tDCS could categorically result in a loss of authenticity amongst chess players. If anything, with better mental stamina, a player can train for longer, think harder and achieve better results — the makings of a more authentic player and self. In terms of the broader regulatory framework, the uncertainties surrounding the therapy/enhancement distinction and the fallacy of the assumption that doping/enhancement is necessarily contrary to authenticity means that this is a troublesome ethical principle for WADA. When placed in the context of the biological differences debate, the argument that doping would result in a loss of authenticity is both manifestly contradictory and hypocritical.

²¹⁵ *ibid.*

²¹⁶ Bostrom and Sandberg (n 7) 321.

²¹⁷ *ibid* 327.

iii) Enhancements such as tDCS can promote the principles of equality and fairness

Regarding the goal of fairness, WADA's Ethics Panel is tasked with ensuring the provision of 'equal opportunity and [the facilitation of] fair competition'²¹⁸ in a world where 'individuals are fundamentally different and not equal, and the circumstances under which athletes might have to train are not the same'.²¹⁹ In short, WADA commits to a two-tiered version of fairness composed of equal opportunity and fair competition. The section below deals with each in turn. The reality, as is suggested, is that both of these goals are, under the current regime, unrealistic.

In terms of equal opportunity, it is apparent that existing inequalities render this goal fanciful. For example, it is readily apparent that athletes from richer countries have better access to (non-illicit) enhancing methods, such as hypoxic tents, better training facilities, better diets and better physiotherapists. Indeed, evidence suggests that sporting success correlates with government funding. Australia devoted \$547 million to sports funding in the years prior to the Athens Olympics — and brought home 17 gold medals.²²⁰ In simple terms, they likely won more because they spent more.²²¹ Deducing that one's opportunity is ultimately defined by which country one represents, Park argues that the use of tDCS is unfair because it helps to further perpetuate this economic inequality 'as only athletes of means would likely benefit from the technique'.²²² Since the notion of 'fairness requires that rich and poor athletes have equal opportunities to win',²²³ it follows that, as Park argues,²²⁴ allowing the use of tDCS is to advantage rich athletes with access to better opportunities, strengthening the argument for a WADA ban. However, this argument is flawed: it fails to appreciate that tDCS devices cost much less on a relative scale as

²¹⁸ 'WADA Ethics Panel: Guiding Values in Sport and Anti-Doping' (WADA, October 2017)

<<https://www.wada-ama.org/en/resources/wada-ethics-panel-guiding-values-sport-and-anti-doping>> accessed 29 April 2022.

²¹⁹ *ibid.*

²²⁰ Savulescu, Foddy and Clayton (n 89) 668.

²²¹ *ibid.*

²²² Kwangho Park, 'Neuro-Doping: The Rise of Another Loophole to Get Around Anti-Doping Policies' (2017) 3 *Cogent Social Sciences*.

²²³ *ibid.* 4.

²²⁴ *ibid.*

compared with other forms of enhancement methods. As Park himself notes,²²⁵ the starting price of these devices for athletes is around \$500. Would this pricing not be more appealing in terms of funding as compared to, say, a \$75,000 Hydroworx pool that Mo Farah used as his underwater treadmill?²²⁶ It may be the case that even \$500 is a high amount for athletes in certain countries, but it certainly goes towards leveling the playing field much more than other more expensive enhancements. Given that wealthier athletes already have access to various expensive performance enhancing devices, and recognising that countries like Australia can afford to (and actually do) fund athletes and produce champions and medals, why is it then that athletes from poorer countries should be estopped from benefiting from access to this relatively cheap form of neuro-enhancement? Admittedly, it is notionally difficult to establish a threshold of money spent to render an unfair advantage, but given that there are permitted devices, such as the Babolat tennis racket capable of recording ball spin and speed that costs close to \$400, the use of tDCS should be within the acceptable range unless ‘one can give a good moral argument for why it is only tDCS, out of all types of equipment, that should be banned’.²²⁷

In the context of chess, a similar story emerges. Chess success correlates with the amount of government funding, with India tipped to be the dominant force in the world of chess in the coming years. In 2019, the Indian government announced an increase in grants from 4 crore to 5.5 crore for development of youngsters and top chess players.²²⁸ Funding is crucial for travelling to tournaments, receiving scholarships and winning prize monies. In countries lacking funding, whether governmental or commercial, the same kind of inequality exists, which means that the introduction of the relatively cheap tDCS may be a tool to equalise the lack of such funding by aspiring chess masters. In turn, the introduction of tDCS does not inhibit fair opportunity but promotes it.

²²⁵ *ibid.*

²²⁶ ‘Cutting Edge Science at Mo Farah’s Oregon Training Camp’ (*British Journal of Sports Medicine*, 20 April 2012) <<https://blogs.bmj.com/bjbm/2012/04/20/cutting-edge-science-at-mo-farahs-oregon-training-camp/>> accessed 29 April 2022.

²²⁷ Petersen and Lippert-Rasmussen (n 14).

²²⁸ Sagar Shah, ‘The Contribution of The Government Towards Indian Chess – Chessbase India’ (*Chessbase India*, 12 August 2019) <<https://chessbase.in/news/The-support-by-Government-of-India-towards-Indian-chess>> accessed 29 April 2022.

Turning now to the fair competition argument, it is argued that this is an unrealistic ideal if fair competition is understood in either the strict sense of absolute equality in competition, or the sense that athletes should not have an unfair advantage. This is in large part because some individuals inevitably benefit from the ‘genetic lottery’. Finnish skier Eero Maentyranta, for example, who won two Olympic gold medals in 1964, enjoyed a genetic mutation ‘naturally’ giving him 40-50% more red blood cells compared to the average.²²⁹ Similarly, Michael Phelps, the most decorated Olympian of all time, enjoys ‘disproportionately long arms and overly lax joints’,²³⁰ giving him a competitive advantage in swimming races. We see this trope appear in the game of chess as well. For example, Howard’s 2011 study suggests that ‘natural talent affects chess expertise development’²³¹ and that ‘chess-prodigies probably do have great natural talent’,²³² with various other studies noting a link between chess skill and IQ.²³³ We do not, however, require chess players to be categorised by IQ points, just like we do not require Phelps to compete in categories with other athletes who have the same wingspan to height ratio. Conversely, what sport competitions do is try to ‘eliminate or compensate for inequalities between individuals and groups upon which they exert little or no control and for which they cannot be held responsible’.²³⁴ To this end, ‘women do not generally compete with men in tennis competitions’²³⁵ or in weightlifting competitions (although one potential exception is of transgender women and intersex athletes, especially after the recent reversal of the Olympic guidelines in relation to transwomen,²³⁶ as well as the approval by IOC to allow transgender athlete

²²⁹ Savulescu, Foddy and Clayton (n 89) 667.

²³⁰ Bostwick and Joyner (n 206).

²³¹ Robert W Howard, ‘Longitudinal Effects of Different Types of Practice on the Development of Chess Expertise’ (2011) 26 *Applied Cognitive Psychology*.

²³² Robert W Howard, ‘Linking Extreme Precocity and Adult Eminence: A Study of Eight Prodigies At International Chess’ (2008) 19 *High Ability Studies*.

²³³ Guillermo Campitelli and Fernand Gobet, ‘Deliberate Practice’ (2011) 20 *Current Directions in Psychological Science*.

²³⁴ Sigmund Loland, ‘Performance-Enhancing Drugs, Sport, and the Ideal of Natural Athletic Performance’ (2018) 18 *The American Journal of Bioethics*.

²³⁵ Jim Parry and Irena Martínková, ‘The Logic of Categorisation in Sport’ (2021) 21 *European Journal of Sport Science* 5.

²³⁶ Sean Ingle, ‘Trans Women Should Not Have to Reduce Testosterone, Say New IOC Guidelines’ (*The Guardian*, 16 November 2021)

<<https://www.theguardian.com/sport/2021/nov/16/trans-women-should-not-have-to-reduce-testosterone-say-new-ioc-guidelines>> accessed 29 April 2022.

Laurel Hubbard to compete at the 2020 Tokyo Olympics),²³⁷ and heavyweight boxers do not fight lightweight competitors. Equipment is standardised and classification systems operate on criteria such as biological sex, body mass and age. Even jockeys in horse races are weighed before every race and water-saddled to equalise weight differences. The concept of equality is clear, transparent and principled. As such, the current sporting world evidently tolerates a degree of ‘unfairness’ and simply seeks, insofar as possible, to reduce its effects. In fact, it could be argued that tDCS, when used in chess competitions, could serve to do just this. On the policy level, Bostrom and Sandberg note that, as it relates to the ‘talent gap, some individuals at ‘the low end of the performance spectrum’²³⁸ have the potential to benefit more than those ‘at the high end whose brains are already functioning close to their biological optimum’.²³⁹ Hence, allowing the use of tDCS would help close the chess skill gap and help achieve the goal of ‘fairness’. Indeed, Tännsjö goes as far as to argue for the idea of ‘biological egalitarianism’, holding that, if we have the opportunity to level out our skills through the use of neuro-enhancements, we should do so.²⁴⁰ He holds that WADA has hitherto taken a ‘fascistoid ideology’ which suggests that those who are naturally stronger should win over those who are weaker and argues that neuro-doping can help ‘liberate elite sport from its perverse view of fairness’.²⁴¹ Surely, on this account, Tännsjö must be correct — it is not consistent to argue that we should have fair competition whilst at the same time maintaining our commitment to the acceptability of not equalising genetic and economic conditions. This is instinctively true: when a chess player plays a wrong move, we do not blame it on the genetic disadvantage of being less intelligent or having a poorer memory or shorter attention span; instead, we regard it as their mistake of miscalculation, a human error attributable to a lack of preparation or a lapse in judgement. If this is true, then surely, where there *is* a genetic disadvantage, there should be a ‘prima facie moral obligation’ to allow the enhancement of disadvantaged individuals,²⁴²

²³⁷ Tariq Panja and Ken Belson, ‘Olympics’ First Openly Transgender Woman Stokes Debate on Fairness’ (*The New York Times*, 31 July 2021) <<https://www.nytimes.com/2021/07/31/sports/laurel-hubbard-trans-weight-lifting.html>> accessed 29 April 2022.

²³⁸ Bostrom and Sandberg (n 7) 329.

²³⁹ *ibid.*

²⁴⁰ Torbjörn Tännsjö, ‘Neuro-Doping as a Means to Avert Fascistoid Ideology In Elite Sport’ (2021) 14 *Neuroethics* 169-178.

²⁴¹ *ibid.*

²⁴² Jeessoo Nam, ‘Biomedical Enhancements As Justice’ (2013) 29 *Bioethics* 126-132.

because the game of chess should not be fundamentally reduced to a battle of genetics.

In this manner, the use of tDCS in chess and other mind sports at the very least does not seem to contravene the fairness of the sport — both in terms of equal opportunity and fair competition. Indeed, it appears that the two-pronged approach by WADA is unrealistic and hard to achieve.

The above Sections on hard work, authenticity and fairness have each shown that the ethical dimensions of the ‘fundamental rationale’ of the spirit of sport embraced in WADC fail to stand up to scrutiny. It has been shown that the use of tDCS does not necessarily remove hard work in chess but may in fact be construed as an example of commitment to the pursuit of excellence. It has also been shown that the authenticity argument is flawed because it is not plausible for WADA to draw a clear line between therapy and enhancement, which means that it is impossible for WADA to devise rules or a set of policies based on this imaginary ‘line’ with any conceivable consistency or accuracy. Furthermore, via tDCS in chess, it was suggested that certain forms of doping may in fact *aid* in the expression of an athlete’s authentic self. Finally, the idea of fairness as a cornerstone of WADC was questioned in light of the inevitable economic, social and genetic inequality that otherwise pervades. It was again proffered that tDCS may go some way to alleviating these inequalities. These conclusions are not to suggest, however, that ethical justifications for regulation are unimportant; the issue here is that the current ethical justifications have difficulty standing on their own merits and, when applied in the context of tDCS and chess, are further challenged and undermined. It is abundantly clear that WADA must replace its current rules with more suitable ones underpinned by sound philosophies capable of withstanding scrutiny.

Indeed, any sound regulatory framework should be ethically and philosophically justified if it is to retain legitimacy. Legitimacy on a global scale is a pressure point for WADA and one that will only be exacerbated by an arbitrary framework. Despite the ratification by 191 countries of the UNESCO International Convention against Doping in Sport, which, in practice, allows

governments ‘to align domestic laws and policies with [WADC]’,²⁴³ WADA continues to be plagued with concerns of conflicts of interest. Half of WADA’s core funding is derived from the IOC at \$34.6 million, while half of the board’s seats, that is ‘six out of 12 seats on the executive board’²⁴⁴ have been conferred on the IOC (with the other half appointed by governments around the world). In turn, WADA regulations are heavily influenced by politics. Notably, the rejection of a WADA proposal to impose a total ban on Russia after revelations of state-concerted efforts of doping due to a critical rejection vote from a member of IOC sent a ‘chilling message’ that WADA is merely a ‘service provider [rather] than a shaper of its policy’.²⁴⁵ Furthermore, WADA does not enjoy close relationships with NADOs, the independent organisations in charge of testing their countries’ athletes. In particular, it is difficult for WADA to effectively monitor NADOs as differing financial constraints have necessitated a ‘vast gap ... between testing programs in various countries’.²⁴⁶ On the other hand, NADOs seldom view WADA positively as its decisions ‘[a]re seen as arbitrary and d[o] not always reflect NADOs’ opinions and priorities’.²⁴⁷ An extensive elaboration of the power relationships is beyond the scope of this article, but it suffices to say that WADA has reached a critical point that, if the ethical and philosophical justifications for its anti-doping regulatory framework are not at least strengthened, their legitimacy would be further questioned. Policies that rest on sound principles provide at least a common basis for countries and organisations alike to cooperate going forward. Any delay in addressing these issues would only set back IOC’s and WADA’s respective attempts to protect their powers and to remain a ‘trust device’.²⁴⁸ This difficulty

²⁴³ ‘Lebanon becomes 191st country to ratify UNESCO Anti-Doping Convention’ (WADA, 2020)

<<https://www.wada-ama.org/en/news/lebanon-becomes-191st-country-ratify-unesco-anti-doping-convention>> accessed 2 October 2022.

²⁴⁴ Eddie Pells, ‘IOC’s Large Role In Anti-Doping Creates Conflict Of Interest’ (*AP NEWS*, 2019)

<<https://apnews.com/article/winter-olympics-ap-top-news-olympic-games-edwin-mose-s-health-596457e4286d4c56a8394881a022665d>> accessed 28 September 2022.

²⁴⁵ *ibid.*

²⁴⁶ *ibid.*

²⁴⁷ Ekain Zubizarreta and Julie Demeslay, ‘Power Relationships Between The WADA And Nados And Their Effects On Anti-Doping’ (2021) 8 *Performance Enhancement & Health*.

²⁴⁸ Fabien Ohl, Lucie Schoch and Bertrand Fincoeur, ‘The Toxic Doxa Of “Clean Sport” And IOC’S And WADA’S Quest For Credibility’ (2020) 56 *International Review for the Sociology of Sport* 1116-1136.

has in part stemmed from the eagerness that IOC and WADA have so far exhibited, causing them to '[trap] themselves within their own staged discourse', 'outbidding promises of clean sport, which turned ineffective and even toxic'.²⁴⁹

In summary, it is crucial that a solid and realistic framework based on sound ethical principles that can appeal to relevant stakeholders is in place. This article will now turn to the final Section to discuss a proposed framework.

III. CONSTRUCTING THE FUTURE - ALTERNATIVE PATHWAYS FOR FUTURE REGULATION

This final Section draws on the lessons learnt from the above discussions and offers suggestions for how WADA's anti-doping regulatory framework should be reformed.

Criticisms against WADA's anti-doping policies and suggestions for amendments have generated a spectrum of academic debate, which this article has attempted to summarise. On one end are those such as Savulescu's, who hold that athletes are 'condemned to cheat' as the goal of 'cleaning up the sport is unattainable'.²⁵⁰ Instead, Savulescu posits that regulation should be focused on the welfare of the athlete, seeing that '[o]ur crusade against drugs in sport has failed. Rather than fearing drugs in sport, we should embrace them'.²⁵¹ On the other end of the spectrum, Sandel argues that allowing enhancements would be tantamount to 'eroding human agency'²⁵² and that our admiration of the excellence of athletic performance would fade because we would come to admire the inventor and not the athlete themselves. However, between these two contrasting positions, a consensus may emerge. First, even advocates against doping admit that the spirit of sport criterion is somewhat of an 'elusive concept'.²⁵³ Secondly, most theorists also appear to agree that the extent of regulation should correlate with the nature of the sport — a theme already explored in Sections 1.3 and 2.1, which argued for different considerations to apply inter-sports (as between physical and mind sports) and intra-sports (as between and within different mind sports).

²⁴⁹ *ibid.*

²⁵⁰ Savulescu, Foddy and Clayton (n 89) 666.

²⁵¹ *ibid* 670.

²⁵² Sandel (n 85).

²⁵³ Stephen Moston, 'The Spirit of Sport?' (2013) 2 Performance Enhancement & Health.

In light of this, the author suggests and supports the following reforms:

- (a) That athletes' health be given priority and the spirit of sport be specifically defined and eventually phased out as a fundamental rationale;
- (b) That there be a shift in focus away from the spirit of sport criterion and towards the enhancement criterion which should be *outcome*-based; and
- (c) That there occur decentralisation of governance away from WADA to sporting governing bodies best placed to appreciate the values and nature of the sport, such that any future regulations can reflect the nuances of that particular sport.

3.1 Spirit of Sport

Regarding the spirit of sport reform, various suggestions have been hitherto proffered. For example, Henne et al.²⁵⁴ argue that the concept should be abandoned entirely, and focus should instead be on health and performance enhancement. Taking a less controversial approach, Loland and McNamee argue that the spirit of sport concept should be reformulated through finding an 'overlapping consensus among all WADA stakeholders'.²⁵⁵ They offer a reformulated understanding of the concept, one that focuses on the 'integrity of sporting competition' whilst balancing the need for athlete protection. Diverse though these views are, they appear to have one common denominator: the health of the athlete.

The preferred approach that this article offers is to incorporate or reshape the 'spirit of sport' into a health-centred concept. As Savulescu et al. suggest, prioritising the welfare of the athlete is the primary concern²⁵⁶ — *if* the

²⁵⁴ Kathryn Henne, Benjamin Koh and Vanessa McDermott, 'Coherence Of Drug Policy In Sports: Illicit Inclusions And Illegal Inconsistencies' (2022) 48-55.

²⁵⁵ Loland and McNamee (n 73) 325.

²⁵⁶ Savulescu, Foddy and Clayton (n 89) 670.

health of an athlete is at risk, *none* of the fundamental underlying principles of hard work, authenticity and fairness, even if shown to be sound, would really matter. The reality is that WADA's regulations over the last 20 years have largely failed to prevent systematic abuse and widespread drug use in sports and, in fact, its strict insistence on a deontological approach can paradoxically lead to proliferation of black markets where 'the safety of the product is questionable',²⁵⁷ worsening the health of athletes. Savulescu et al. raise an analogy between the prohibition of doping and that of the prohibition of alcohol in America which, rather than reducing drinking, increased the volume of alcohol consumption.²⁵⁸ The same logic is applicable here: despite WADA's strict, duty-based and over-inclusive form of regulation, empirical evidence has shown that the demand for enhancement by athletes has not decreased. The emphasis placed on health is not to suggest that WADA should take a completely hands-off approach and allow all enhancing substances/methods; rather, a more measured and considered approach that places health at the centre should be adopted. This would include conducting risk analyses on substances/methods which are heavily abused but have insignificant adverse effects and allowing their use within the parameters of health, or an implementation of a threshold system which would allow for some degree of intake/intervention provided the line is not crossed and the athlete's health is not compromised.

In other words, a reformulation of the spirit of sport to a health-centred concept would also mean that the discourse is driven by science. By focusing on methods/substances that *actually* enhance (see below), this would encourage greater scrutiny by scientists and experts and generate a more transparent and open system. In this manner, defining 'doping' would be brought back to the centre-stage with the result that the discourse has become more focused and less subjective.²⁵⁹

Regardless of how the conception of health would be incorporated into a newly reformed understanding of the spirit of sport, WADA must alleviate the burden it places on the interpretation of the spirit of sport. This should be achieved over time and not in one sweeping change that uproots what

²⁵⁷ *ibid* 669.

²⁵⁸ *ibid* 670.

²⁵⁹ See also de Hon (n 49) 672.

has, until now, been the entire basis of regulation. A more effective course of action is for the spirit of sport both as a criterion and a concept to fade away *gradually* to become an ideal and an aspiration, much like the spirit of Olympism, rather than being relied on for regulatory purposes. As part of this ‘phasing-out’, the emphasis of regulation should be shifted to the enhancement criterion — the second suggestion to which we shall now turn.

3.2 Enhancement Criterion

At its core, what this proposal does is suggest a shift in attitude of WADA from the policing of behaviour or enforcing the duty of athletes not to dope to one based on the *actual outcome* or *actual enhancement*. In other words, a shift to a results-based system is proposed, where the burden would be a proof of enhancement, not the presumption of it.

Substantively, this would mean removing the words ‘has the *potential* to enhance’ from Article 4.3.1.1 so that it reads:

‘...with other substances or methods, ~~has the potential to enhance or~~ enhances sport performance’;

and further removing the words ‘or potential’ from 4.3.1.2 so that it reads:

‘that the use of the substance or method represents an actual ~~or potential~~ health risk to the athlete’.

In this manner, the enhancement criterion would only ban substances and techniques that *actually enhance* or improve performance in a measurable manner. This would be emblematic of a shift in thinking towards a consequence-based framework that places emphasis on the practical realities of doping, such as recognising that cognitive enhancement via tDCS is unlikely to make a difference between winning and losing in the context of chess. It might be argued that this would fail to capture those ‘ill-intending’ athletes who attempt to gain an advantage through doping but turn out unsuccessful in enhancing their performance. Two responses are offered here. First, an attempt to cheat is, of course, potential evidence of the duplicitous intent of the athlete. However, where this attempt *fails*, it is not only difficult to prove there *was* an attempt, but it is also even more difficult to prove the presence of intent.

Attempting to decipher intention in every situation is costly, time-consuming and speculative. Secondly, it is arguably not much of a problem if these athletes remain unpunished. If WADA is truly concerned about clean sports and ‘playing true’, then, surely, an athlete who gains no advantage, despite their intention to, does not affect competition in any way. If such an athlete does gain *actual* (the meaning of which is explained below) enhancement, it would not be too late to administer punishment. The remaining issue with this shift is that the current definition stipulated by WADA’s enhancement criteria, as it stands, is insufficient: it does not give guidance as to *how much*²⁶⁰ or *what type* of enhancement is ‘too much’. For this, we turn to the argument in favour of decentralisation.

3.3 Decentralisation

It is proposed that any future change in WADA regulation should be attuned to the fact that different sports entail different qualities and that ‘each sport [should] determine its own rules’ and ‘whether neuro-doping poses a risk to its ethos’.²⁶¹ In this light, what is meant by *actual* enhancement may differ from one sport to another. This article does not advocate that athletes be tested on an individual level for potential enhanced performance but instead regulation should, as far as possible, give power to individual sporting bodies themselves to determine whether a substance or technique violates the ‘spirit’ of a particular sport and hence *actually* enhances in that specific context. For example, it may be that FIDE ultimately decides that tDCS should not be banned in chess because of a lack of evidence of actual enhancement in chess performance, and that it does not violate the spirit and values of chess. On the other hand, the World Athletics may decide, taking into account similar considerations, that, on balance, tDCS does enhance and violate the spirit of long-distance running and therefore institute a ban. Further nuances may be added to this argument, such as allowing these sporting bodies to determine the degree of enhancement allowed, or to limit the ban to certain categories, such as, for example, only banning tDCS for track races above 1500m. This approach would not only free up WADA’s resources but allow for a more measured policy that truly reflects ‘respect for self and other participants’²⁶² and the physical/cognitive divide.

²⁶⁰ Except with respect to 4 exceptions in Category S3 in PLIS.

²⁶¹ Davis (n 13) 651.

²⁶² WADC (n 45).

This argument for decentralisation is evidently not without its flaws. Issues of individual sports bodies' funding, resources and accountability may arise — having a centralised arbiter, WADA, reduces the chances of a conflict or inconsistencies between individual sporting bodies, and helps (at least ostensibly) retain impartiality. Nevertheless, it is conceivable that WADA could still play a similar role of ultimate oversight in this decentralised model, albeit with greater delegation powers to individual sporting bodies such as the World Athletics and FIDE. Further, this arrangement could help alleviate WADA's legitimacy problems. Devolving decisions in order to show more realism and consistency could serve to boost WADA's approval amongst separate sporting federations and indeed amongst professional athletes.

A further argument against this model is that it could potentially run into problems with inconsistencies. For example, caffeine may become a banned substance in chess, but not in long-distance running, which may become confusing for spectators and athletes alike. Despite this, a tailored model would allow more room for just results and would solve the fundamental issues of acceptable enhancement and the lack of a universally understood concept of the ostensible spirit of sport. Such a tailored approach could also prevent recurrence of the Raducan-type controversy, as the gymnastics governing body would have regard, unlike WADA, to the acceptability of the common cold pill *in the context of gymnastics as a sport (and its core values)* and specifically to *how* and to *what extent* it had aided, if at all, Raducan's ability to perform complicated somersaults with such accomplished grace and precision. Moreover, the advantage of decentralisation lies in reaction time. When a new technique, such as tDCS, emerges or when a scientific advancement reveals fresh findings, sporting bodies can engage with analysis of the risks associated with a *particular sport* in the context of the particular nature of that sport, its core values and the degree of impact on those values.

In terms of implementation, any such proposal would, of course, require consent and active agreement amongst all stakeholders, including IOC, governments, individual sporting bodies and athletes. It is difficult to say for certain whether such a model would be welcomed with open arms by all parties, especially given the complicated political interplay between IOC, WADA and national governments. Nevertheless, it is argued that this model would work

best in the age of emerging NIBS technologies such as tDCS, which have the potential to enhance both physical and mind sports performance in both cognitive and physical aspects — a lacuna which WADA can no longer afford to neglect.

CONCLUSION

This article has argued, through the prism of chess and tDCS, that WADA's anti-doping regulations, in particular the enhancement criterion, are unclear and over-inclusive, yet the concept of cognitive enhancement is underexplored, and that, further, the fundamental principles underlying these regulations, particularly that of the spirit of sport, are themselves intrinsically flawed. Emerging NIBS technologies such as tDCS highlight loopholes in the current anti-doping regime that fails to pay regard to the existence of the physical and mental divide, the nature of cognitive enhancement and the nature of each sport and its core values. Finally, recommendations in favour of a shift to a decentralised, consequence-based regulatory framework combined with a reformulated conception of the spirit of sport are offered, specifically, the reformulations of Articles 4.3.1.1 and 4.3.1.2 of WADC so as to require *actual* enhancement rather than *potential* enhancement. It is proposed that such a devolved regime would relegate more power to individual sporting bodies. In such a manner, athletes in both the mental and physical sporting realms could genuinely be regarded to 'play true'.

WINTER ISSUE



2023

International Corporate Liability: The Past, Present, and Future of ‘Class Action Tourism’ in England

Patrick Kenny*

ABSTRACT

Class action tourism refers to claims which are brought by a large number of claimants against parent companies in jurisdictions that are perceived to be claimant-friendly, and which are based on alleged wrongdoing by those companies’ subsidiaries elsewhere. Such claims, commonly arising under tort, environmental, or human rights law, have become increasingly prominent in recent years. This article explains the background to this trend before discussing two important matters in detail: parental liability and issues of jurisdiction. By analysing these two issues, it seeks to explain both the recent prominence of these claims and their prospects in coming years. In particular, it argues that political and legislative developments are likely to be crucial to the future of class action tourism. On the one hand, more stringent requirements upon corporations with regard to human rights and sustainability, combined with recent judgments, suggest that the courts may be increasingly willing to rule in favour of the claimants on the issue of parental liability. On the other hand, the EU legal regime aided the development of these cases and, post-Brexit, English courts have greater flexibility to decline jurisdiction over them. As a result of these developments, more cases may be struck out or stayed for jurisdictional reasons, but those that survive any jurisdictional challenge are less likely to be struck out on the ground that they cannot demonstrate potential parental liability and may be more likely to achieve ultimate success at trial or via settlement.

*LLM (BPP) ‘23. BA (Oxon) Arabic and Islamic Studies ‘18. Solicitor specialising in environmental and public law. Views are my own.

INTRODUCTION

Class action tourism has become increasingly prominent over recent years, both in English courts and abroad.¹ Such cases often involve claims for substantial damages, a very large number of claimants, and extensive media coverage.² Recent judgments exemplifying such claims have been discussed in the national and legal press: *Okpabi v Royal Dutch Shell Plc*,³ *Município De Mariana v BHP Group (UK) Ltd*,⁴ and *Vedanta Resources Plc v Lungowe*.⁵ These, and others, are examined below.⁶ An archetypal case would involve a subsidiary in the energy or mining sector allegedly allowing pollution or poor working conditions that harm local people in the Global South; all of *Okpabi*, *Município De Mariana*, and *Vedanta* fit this model. A claim is brought elsewhere, often in Europe, the USA, or Australia, against a parent company incorporated there (in addition to the immediate subsidiary) on the basis that the parent company is liable.

¹ For ease, I use ‘England’ and ‘English’ when the jurisdiction or court is that of England and Wales.

² The claim against BHP Group (UK) Ltd, discussed in more detail below, for example, involved over 200,000 claimants, a claim for billions of Brazilian reais in damages, and was reported in media sources including the Financial Times, Bloomberg, and Reuters; see *Município De Mariana v BHP Group (UK) Ltd* [2022] EWCA Civ 951, [2022] 7 WLUK 75 [2] and [14] and Jane Croft, Neil Hume and Michael Pooler, ‘BHP will face £5bn High Court lawsuit over Brazilian dam disaster’ (2022) Financial Times <<https://www.ft.com/content/75ffb4eb-195f-460c-b2b0-dfeca3b60ab9>> accessed 14 September 2022, Jonathan Browning and Katharine Gemmell, ‘BHP Set to Face \$7 Billion UK Suit Over Deadly Dam Collapse’ (2022) Bloomberg UK <<https://www.bloomberg.com/news/articles/2022-07-08/bhp-set-to-face-7-billion-london-suit-over-deadly-dam-collapse?leadSource=uverify%20wall>> accessed 14 September 2022, and Kirstin Ridley, ‘London judges reverse course to reopen \$7 bln Brazil dam lawsuit against BHP’ (Reuters, 2021) <<https://www.reuters.com/business/energy/london-judges-re-open-69-bln-brazil-dam-lawsuit-against-bhp-2021-07-27/>> accessed 14 September 2022.

³ *Okpabi v Royal Dutch Shell Plc* [2021] UKSC 3, [2021] 1 WLR 1294.

⁴ *Município De Mariana v BHP Group (UK) Ltd* [2022] EWCA Civ 951, [2022] 7 WLUK 75.

⁵ *Vedanta Resources Plc v Lungowe* [2019] UKSC 20, [2020] AC 1045.

⁶ For such an important field, there may appear to be relatively few actual judgments or trials; this is in part because many cases are settled, such settlement being a result of the various factors and interlocutory decisions detailed in this article.

Various requirements have to be satisfied for class action tourism claims to be brought successfully in England. It is trends or developments concerning these factors that have resulted in the current prominence of this form of litigation, and that are likely to influence its future. These requirements, inherent in this sort of claim, demonstrated by the cases that have been brought and which are detailed in this article, are: (1) a corporate structure involving English-domiciled parent companies with subsidiaries abroad; (2) wrongdoing committed by a subsidiary; (3) losses resulting from that wrongdoing;⁷ (4) a relevant cause of action under foreign or English law; (5) a claim brought against the parent and subsidiary in England is more attractive and viable than bringing the claim elsewhere (because, for instance, the English courts and judiciary are held in high regard, English lawyers are increasingly experienced with these sorts of claim, better funding mechanisms may be available in England, and greater media coverage in England can assist in achieving settlement);⁸ (6) English courts have jurisdiction to hear such cases; and (7) English courts are willing to find the parent companies liable. The requirements for this last condition may vary but often involve the usual tort elements including loss and causation. A particularly important issue here involves the need for a sufficient link between the parent and the wrongdoing to find liability on the part of the parent company.

The last two factors, those of liability and jurisdiction, are especially influential and have seen significant developments over recent years. They are likely to see further developments in the near future. This article looks at these two factors in turn. It first explains how parental liability has historically been treated by the English courts and the significant recent decisions that the Supreme Court has made in this regard. It then explores what we might expect in the future, arguing that developments in legislation and corporate practice may prove very helpful for claimants bringing foreign class action claims. The second half of the article examines the issue of jurisdiction. It first explains the position taken by the courts over recent years and why this position has made it relatively easy for claimants to bring their claims in England. It then again argues that political and legislative developments, particularly as a result of Brexit, will have a crucial effect on these cases, which is likely to be detrimental to those

⁷ Note that, for group claims, the claimants must all have a sufficiently similar interest.

⁸ Damian Grave, Maura McIntosh and Gregg Rowan (eds), *Class Actions in England and Wales* (1st edn, Street & Maxwell 2018) para 10-053.

seeking to bring their claims in England. It suggests that, in coming years, these cases will be far more dependent on access to justice in the jurisdiction where the alleged wrongdoing occurred and will be more likely to fail on jurisdictional grounds.

This may result in more cases being struck out at an early stage as a result of jurisdictional challenges. Those that are not will be less likely to be struck out or fail on the issue of whether a duty of care exists between the parent company and the claimants (or some statutory basis for parental liability applies). This latter issue has often been seen as particularly difficult for claimants to demonstrate. While success here does not mean that a positive outcome in any final judgment or via settlement is guaranteed, it may increase the probability of such an outcome.

I. PARENTAL LIABILITY

A. Early developments in case law

Generally, in these cases, both the parent company and the subsidiary are defendants. This section focuses on parental liability because: (1) this has tended to be a more complicated issue; (2) there have been important developments in the case law regarding parental liability over recent years; and (3) persuading the court of — at least the potential — liability of the parent company has been vital for jurisdictional reasons, as detailed below.

Liability is generally claimed under tort law and sometimes under specific statutory regimes in the jurisdiction where the wrongdoing took place. Only the former is discussed here in order to arrive at an analysis of general application: the same or similar tort principles apply across cases brought in England, whereas statutory liability will depend on the specific legislation in question and will, therefore, differ between cases. Such tort claims are generally based on local law. This law, however, is often considered sufficiently close to English law that English case law and principles apply. This, for instance, was the case for Zambian tort law in *Vedanta* and Kenyan tort law in *AAA v Unilever Plc*.⁹

⁹ *AAA v Unilever Plc* [2017] EWHC 371, [2017] 2 WLUK 681.

The crucial question in the context of tort claims, and the one examined here, is whether a duty of care arises on the part of the parent company. This is complex and case specific, but the analysis below explains important recent developments in England and principles to apply when evaluating a claim's chances of success or failure.

Generally, English law avoids piercing the corporate veil, drawing a distinction between the company (as a separate legal entity) and its members. Shareholders (who might themselves be corporate entities) are thus insulated from corporate liability where a claim is brought against the company they own.¹⁰ However, liability on the part of shareholders can be found where a duty of care arises independently on the part of the parent company towards the aggrieved party. Liability therefore arises not because the parent is the shareholder of a liable company; rather, the parent is liable for its own actions towards the claimants.¹¹

In the 1990s, the courts held in several personal injury claims (at pre-trial stage) that a parent may owe a duty of care to a subsidiary's employees.¹² This area of law continued to develop, and a particularly crucial judgment handed down after a full trial and frequently cited in foreign class action claims, came in the case of *Chandler v Cape Plc*.¹³ *Chandler* involved an individual employed by a subsidiary of Cape Plc, who was exposed to asbestos during his employment and was subsequently (after the dissolution of the subsidiary) diagnosed with asbestosis. He then brought a claim against Cape Plc alleging negligence.¹⁴ In *Chandler*, Arden LJ set out circumstances which may result in parental responsibility: (1) both parent company and subsidiary are involved in sufficiently similar businesses; (2) the parent has, or should have, superior health and safety knowledge; (3) the parent knew, or ought to have known, the subsidiary's system of work was unsafe; and (4) the parent knew or

¹⁰ *Salomon v Salomon* [1896] UKHL 1, [1897] AC 22.

¹¹ James Goudkamp and Donal Nolan, *Winfield & Jolowicz on Tort* (20th edn, Sweet & Maxwell 2020) para 5-037.

¹² See, for example, *Lubbe v Cape Plc* [1998] EWCA Civ 1351, [1998] CLC 1559.

¹³ *Chandler v Cape Plc* [2012] EWCA Civ 525, [2012] 1 WLR 3111.

¹⁴ *ibid* [1].

ought to have known that 'the subsidiary or its employees would rely on its using that superior knowledge for the employees' protection'.¹⁵

Notably, *Chandler* involved liability for employees that were not third parties, loss as a result of asbestosis (a type of loss that courts and statute have been sympathetic to), and loss due to the fact that the subsidiary no longer existed.¹⁶ However, further developments have occurred in the field of class action tourism in addition to these factors. These cases have primarily concerned the interlocutory stage but, as explained further below, such developments are nevertheless significant. In particular, after various judgments in which the lower courts had refused to countenance parental liability, the Supreme Court handed down two significant pre-trial judgments that favour claimants in class action tourism cases.

B. Vedanta

The first was *Vedanta*, involving a claim against the UK-domiciled Vedanta Resources Plc ('**VR**') and their Zambian mining subsidiary, Konkola Copper Mines Plc ('**KCM**'), that alleged losses from toxic discharge in Zambia. In *Vedanta*, the court set out that: (1) at the pre-trial stage, the claimants only needed to prove that, based on the facts available, their claim against the parent was arguable to proceed;¹⁷ (2) the relevant category of common law negligence was not new, and the test from *Caparo v Dickman*¹⁸ (a three-stage test that is applied to determine whether a duty of care exists in novel scenarios) was unnecessary;¹⁹ and (3) the crucial question was how 'the parent availed itself of the opportunity to take over, intervene in, control, supervise or advise the management of the relevant operations ... of the subsidiary'²⁰

The court said it was not possible to list prescriptively the categories or models that could give rise to parental liability, but did indicate what it found

¹⁵ *ibid* [80].

¹⁶ Harry Steinberg, Michael Rawlinson and James Beeton, *Asbestos: Law & Litigation* (1st edn, Sweet & Maxwell 2019) para 4-014.

¹⁷ *Vedanta* (n 5) [61].

¹⁸ *Caparo v Dickman* [1990] 2 AC 605.

¹⁹ *Vedanta* (n 5) [49], [56].

²⁰ *ibid* [49].

suggestive in *Vedanta*, including: (1) evidence that VR took, or even just claimed to take, ‘responsibility for the maintenance of proper standards of environmental control over the activities of its subsidiaries’;²¹ and (2) evidence that VR was involved in training, monitoring, and enforcing standards (although Lord Briggs suggested that producing deficient standards alone could also give rise to a duty of care).²²

The lower courts also remarked upon a management services agreement between defendants that suggested VR was involved in the mine’s management, although the Supreme Court found that less compelling.²³

C. *Okpabi*

In the second case, *Okpabi*, thousands of Nigerian citizens brought claims against a Nigerian subsidiary (‘SPDC’) of the UK-domiciled Royal Dutch Shell (‘RDS’) and against RDS itself, alleging losses caused by oil spills from SPDC-operated pipelines in Nigeria. The judgment was similar to that of *Vedanta*, and offered further guidance: (1) the focus at interlocutory stage should be on whether the case is ‘arguable’ and that ‘factual averments made in support of the claim should be accepted unless, exceptionally, they are demonstrably untrue or unsupportable’;²⁴ (2) the relevant question was ‘the extent to which the parent did take over or share ... the management of the relevant activity’;²⁵ and (3) the Court of Appeal had erred, including by asserting that group-wide policies could never result in a duty of care.²⁶

The claimants in *Okpabi* listed ‘routes’ (for which various pieces of evidence were provided) that they claimed gave rise to liability based on the findings in *Vedanta*, and the Supreme Court in *Okpabi* accepted these to be appropriate headings, though it was stressed that they were not the only issues that could be relevant. These were: (1) RDS controlling the management of

²¹ *ibid* [61].

²² *ibid* [52], [53].

²³ *ibid* [61]; *Lungowe v Vedanta Resources Plc* [2017] EWCA Civ 1528, [2018] 1 WLR 3575 [84].

²⁴ *Okpabi* (n 3) [107].

²⁵ *ibid* [147].

²⁶ *Okpabi v Royal Dutch Shell Plc* [2018] EWCA Civ 191 [2018] Bus LR 1022 [140]; *Okpabi* (n 3) [143].

SPDC's activities; (2) RDS promulgating defective group-wide safety policies which were implemented by SPDC; (3) RDS actively engaging with that implementation; and (4) RDS holding out that it supervised and controlled SPDS's activities.²⁷

Further, and 'significant[ly]' for subsidiary employees,²⁸ the Supreme Court agreed with Sales LJ, who dissented from the Court of Appeal's judgment, that the Shell group's organisation along 'Business and Functional lines' with a 'vertical structure' and 'significant delegation' was suggestive of parental control and thus liability.²⁹

D. The current position in case law

The cases discussed above are generally encouraging for foreign action litigants. They demonstrate the courts' willingness to find that cases against UK-domiciled parent companies are arguable, including on the difficult issue of parental liability. Further, since liability does not revolve around a shareholding relationship, affiliate companies and perhaps even related third parties, such as those financing operations, could be similarly held liable. It has been suggested that *Chandler* could allow for liability on the part of long-term supply contract counterparties, and Sales LJ in *AAA* noted that a consultant's liability could be considered similar to that of the parent company.³⁰ Indeed, such supply chain or outsourcing-related cases have been brought in recent years and, at the very least, *Vedanta* and *Okpabi* are useful for claimants seeking to defend strike out applications, who can plead that their claims are at least arguable and their factual averments should be accepted at interlocutory stage.³¹ To the extent that this grants those who have suffered losses the opportunity to argue their case

²⁷ *ibid* [26].

²⁸ Andrew Tettenborn, *Clerk & Lindsell on Torts* (23rd edn incorporating first supplement, Sweet & Maxwell 2021) para 12-011.

²⁹ *Okpabi* (n 3) [156].

³⁰ Andrew Sanger, 'Crossing the Corporate Veil: The Duty of Care Owed by a Parent Company to the Employees of its Subsidiary' (2012) 71(3) CLJ 478, 480; *AAA v Unilever Plc* [2018] EWCA Civ 1532, [2018] BCC 959 [36].

³¹ See *Hamida Begum v Maran (UK) Limited* [2021] EWCA Civ 326, [2021] 3 WLUK 162 which includes reference to these points at paras [22] and [119], and *Josija v British American Tobacco Plc* [2021] EWHC 1743 (QB), [2021] 6 WLUK 369 which deals with liability across a supply chain.

for redress in a full trial and avoids extended and resource-intensive hearings on preliminary matters, this seems a positive development.

The significant judgments mentioned above took place at the interlocutory stage. It will be a different matter for claimants to show liability at a full trial. A number of the judgments referenced here explicitly recognise that the claimants will have to demonstrate more than just that their case is arguable. Further, success on the issue of parental duty of care does not guarantee the overall success of a claim or that it will not be struck out or rejected on some other basis. Other issues, particularly those typical of mass claims with vast numbers of (often poorly resourced) claimants based overseas (sometimes in dangerous or inaccessible locations), can still present a significant challenge to class action tourism claimants. This can happen at an interlocutory stage and when making an application for a group litigation order or representative action. This includes providing evidence to show that claimants have a sufficiently similar interest.³²

However, the findings from *Okpabi* and *Vedanta* detailed above: (1) demonstrate at least some willingness to countenance parental liability, which has been a particularly difficult issue for claimants; (2) make it harder to strike out these claims on the basis that there is no arguable case; and (3) may encourage defendants to settle. The parties in *Vedanta*, for example, have now settled.³³

It should be noted for completeness that the cases and analysis above relate to scenarios where damage is (allegedly) suffered as a result of poor working practices or environmental causes. A slightly different category of cases involves harm inflicted by unrelated third parties, where an employer or employer's parent company is alleged to be liable for not preventing that harm. Such issues may continue to face difficulties in demonstrating an arguable claim but are beyond the scope of this article.³⁴

³² See, for example, *Jalla v Shell International Trading* [2021] EWCA Civ 1389, [2021] 9 WLUK 299 and *Alame v Royal Dutch Shell Plc* [2022] EWHC 989 (TCC), [2022] 4 WLUK 319, the latter of which is related to the proceedings in *Okpabi*.

³³ Reuters, 'Vedanta Resources settles Zambia copper mine pollution claim' (Reuters, 2021) <<https://www.reuters.com/article/us-zambia-mining-vedanta-idUSKBN2901EL>> accessed 13 February 2023.

³⁴ See, for example, *Kalma v African Minerals Ltd* [2020] EWCA Civ 144 and *AAA* (n 30).

E. Socio-political and legislative developments in the UK

In response to these Supreme Court judgments, it has been suggested that companies could or should seek to avoid liability by reducing oversight of their subsidiaries and maintaining clear divisions between the members of their corporate groups.³⁵ While such an approach might be an ostensibly logical conclusion that can be drawn from an examination of these cases, and while this may be possible in certain circumstances and for certain companies, such an approach does not recognise political and legislative developments over recent years, as well as their likely future trajectory. Both legislative requirements and market practice suggest that such an approach will be difficult to achieve and that claimants in foreign-brought class action cases may have a greater chance of succeeding on the issue of parental liability in the future.

For instance, the Modern Slavery Act 2015 in the UK currently requires sufficiently large companies (which tend to feature more prominently in class action tourism) to disclose what steps they have taken to ensure their operations and supply chains are free from slavery.³⁶ Admittedly, studies suggest that only 88% of companies are complying with their reporting obligations and that statements are not always clear or comprehensible.³⁷ However, the government made various assurances throughout 2020 and 2021 that it would reform and improve the Modern Slavery Act 2015, including through the introduction of fines for those who do not comply with their obligations under

³⁵ For example, Charles Mayo, Chris Owen, Adam Bristow, 'Parent Company Liability' (2019) 2 Ethical Boardroom
<<https://files.simmons-simmons.com/api/get-asset/EBSumemr2019SimonsSimmonsSingles.pdf?id=b1t9b61f5b7cec8259d>> accessed 5 August 2022; Elvis Ojeda, 'Transnational Corporate Liability Litigation and Access to Environmental Justice: The Vedanta v Lungowe Case' (2021) LSE Law Review 6(3) 223, 246.

³⁶ Modern Slavery Act 2015, s 54.

³⁷ FRC, Lancaster University, and Independent Anti-Slavery Commissioner, *Modern Slavery Reporting Practices in the UK: Evidence from Modern Slavery Statements and Annual Reports* (2022) 3
<<https://www.frc.org.uk/getattachment/77c053d9-fe30-42c6-8236-d9821c8a1e2b/FRC-Modern-Slavery-Reporting-Practices-in-the-UK-2022.pdf>> accessed 13 September 2022.

the act.³⁸ Subsequently, in the Queen's Speech of May 2022, a new bill was announced, the provisions of which include a requirement to report on certain key criteria against modern slavery, the creation of a government registry on which all modern slavery statements will be published, and the creation of civil penalties for non-compliance.³⁹ However, the UK government has subsequently changed twice and it remains to be seen whether any such reforms will be made. It appears unlikely that they will happen imminently, given the absence of any commentary on reform from the current UK government.

Another example of a law in the UK that may impact companies seeking to distance themselves from their subsidiaries' actions is the Bribery Act 2010. It incentivises companies to institute procedures against bribery to avoid liability if employees, agents, or subsidiaries engage in corrupt practices, including in overseas operations.⁴⁰ The extent of enforcement is sometimes criticised, but an OECD Report notes that the Act has 'prompted substantial progress' on anti-corruption measures in the UK.⁴¹ This legislation is not new. However, the point here is that, even where a class action tourism claim does not specifically relate to wrongdoing associated with modern slavery or bribery, English parent companies seeking to separate themselves from any knowledge or involvement in their subsidiaries operating abroad in high-risk industries or jurisdictions are likely to find such an approach difficult due to the requirements of these Acts.

Additionally, prospective legislation in the UK seems likely to require continued involvement in and oversight of overseas subsidiaries. One example that might apply to companies operating in jurisdictions from which class action

³⁸ Victoria Atkins, *New tough measures to tackle modern slavery in supply chains* (2020) <<https://www.gov.uk/government/news/new-tough-measures-to-tackle-modern-slavery-in-supply-chains>> accessed 13 September 2022; Dominic Raab, *Human rights violations in Xinjiang and the government's response: Foreign Secretary's statement* (2021) <<https://www.gov.uk/government/speeches/foreign-secretary-on-the-situation-in-xinjiang-and-the-governments-response>> accessed 13 September 2022.

³⁹ Queen's Speech 2022: background briefing notes (2022) 83, 84 <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1074113/Lobby_Pack_10_May_2022.pdf> accessed 13 September 2022.

⁴⁰ Bribery Act 2010, s 7.

⁴¹ OECD, *Implementing the OECD Anti-Bribery Convention* (2017) 78 <<https://www.oecd.org/corruption/anti-bribery/UK-Phase-4-Report-ENG.pdf>> accessed 5 August 2022.

tourism claims have previously emerged (such as Brazil) is the requirement under the Environment Act 2021 to avoid the use of products linked to illegal deforestation and to carry out due diligence to ensure supply chains are free of such products (the exact details of which are still to be outlined in regulations).⁴² Similarly, the government has proposed new sustainability disclosure requirements that would require companies to make assessments of their sustainability impacts, though the details are yet to be finalised.⁴³ Again, this suggests that it is likely to become harder for large multinational corporations subject to these new requirements to distance themselves from involvement in the activities of their subsidiaries abroad.

Moreover, other legislative and regulatory developments have been proposed which, if implemented in the coming years, could have a significant effect on class action tourism cases. For example, a 2017 Parliamentary Report recommended that the UK should impose a duty on companies to prevent human rights abuses by implementing due diligence processes, with failure to do so constituting an offence.⁴⁴ The government's response at the time was to state that it had 'no immediate plans to legislate in this area'.⁴⁵ However, in 2020, an independent research centre recommended legislating to penalise a company's failure to prevent human rights abuses, including environmental abuses. As with the Bribery Act, companies could avoid liability if they had reasonable procedures in place to prevent such abuses.⁴⁶ A similar proposal to this has now

⁴² Environment Act 2021, Schedule 17.

⁴³ HM Government, *Greening Finance: A Roadmap to Sustainable Investing* (2021) 18 <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1031805/CCS0821102722-006_Green_Finance_Paper_2021_v6_Web_Accessible.pdf> accessed 14 September 2022.

⁴⁴ Joint Committee on Human Rights, *Human Rights and Business 2017: Promoting responsibility and ensuring accountability* (HL 153 HC 443) 59 <<https://publications.parliament.uk/pa/jt201617/jtselect/jtrights/443/443.pdf>> accessed 5 August 2022.

⁴⁵ Joint Committee on Human Rights, *Human Rights and Business 2017: Promoting responsibility and ensuring accountability: Government Response to the Committee's Sixth Report of Session 2016-2017* (HC 686) 15 <<https://publications.parliament.uk/pa/jt201719/jtselect/jtrights/686/686.pdf>> accessed 5 August 2022.

⁴⁶ British Institute of International and Comparative Law, *A UK Failure to Prevent Mechanism for Corporate Human Rights Harms* <https://www.biicl.org/documents/84_failure_to_prevent_final_10_feb.pdf> accessed 5 August 2022.

been included in the Options Paper produced by the Law Commission in June 2022, in response to the government's request for reform suggestions to corporate criminal liability, which the government is now to review.⁴⁷ There is no guarantee that these proposals will be transformed into law, but it does suggest a trajectory of development that, as explained below, could have significant impacts on class action tourism cases.

F. Socio-political and legislative developments outside the UK

In addition to these legislative proposals in the UK, comparable legal developments have occurred elsewhere, particularly in Europe. Some of these are likely to apply to English companies if they also operate in the relevant European jurisdiction. These include: (1) Norway's 2021 Transparency Act which requires all companies which do business in Norway, and which pass relatively low employee, balance sheet, or sales revenue thresholds, to carry out and report human rights due diligence and respond to third party requests for information regarding human rights impacts;⁴⁸ and (2) the Netherlands' 2019 Child Labour Due Diligence Act which obliges all companies who supply goods or services to Dutch customers, irrespective of size, to investigate whether their goods or services are the product of child labour and to prepare an action plan if there is a reasonable suspicion of this, for fear of both fines and criminal sanctions.⁴⁹ As a final example, albeit only of a draft instrument, the EU released in February 2022 a draft regulation that would require human rights due diligence be carried out and introduces penalties for those companies — including those which are not based in the EU but have sufficiently high numbers of employees and turnover in the EU — that fail to do so.⁵⁰

Other similar pieces of legislation will not necessarily impose human rights due diligence obligations on English companies directly, but might apply

⁴⁷ Law Commission, *Corporate Criminal Liability: an options paper* (2022) 114 <https://s3-eu-west-2.amazonaws.com/lawcom-prod-storage-11jsxou24uy7q/uploads/2022/06/Corporate-Criminal-Liability-Options-Paper_LC.pdf> accessed 13 September 2022.

⁴⁸ Transparency Act 2021 (*Lov om virksomheters åpenhet og arbeid med grunnleggende menneskerettigheter og anstendige arbeidsforhold (åpenbetsloven)*, LOV-2021-06-18-99).

⁴⁹ Child Labour Due Diligence Law 2019 (*Wet zorgplicht kinderarbeid*, *Staatsblad* 2019, 401).

⁵⁰ European Commission, Proposal for a Directive of the European Parliament and of the Council on Corporate Sustainability Due Diligence and amending Directive (EU) 2019/1937, COM/2022/71 final.

to subsidiaries of English companies in the relevant jurisdiction or apply indirectly where an English company contracts with a company on which the obligations apply directly. For example, France's 2017 Corporate Duty of Vigilance Law and Switzerland's legislation covering child labour and conflict minerals apply to sufficiently large French and Swiss companies respectively, albeit with some exceptions.⁵¹ Germany's 2021 Due Diligence in the Supply Chain Act is similar: it applies to German companies that employ 3,000 employees, though in 2024 this will drop to 1,000 employees, and it introduces mandatory due diligence and reporting requirements that allows for fines to be issued for non-compliance.⁵² Again, the result of this is that English parent companies to which these laws apply, or which have subsidiaries or contractual counterparties to which these laws apply, may struggle to demonstrate to the courts (were a class action tourism claim be brought against them) that they are sufficiently separate and uninvolved in their subsidiaries' conduct to avoid a finding of liability.

A number of international bodies have introduced initiatives encouraging increased care on the part of companies regarding human rights and the environment. The United Nations Guiding Principles on Human Rights, which expect companies to take appropriate steps to determine whether they, their subsidiaries, and their suppliers are respecting human rights, have seen 'swift and widespread' uptake, and a large number of companies have publicly committed to them.⁵³ A similar framework is reflected in the OECD Guidelines for Multinational Enterprises which propose principles for responsible business conduct in areas including human rights, the environment,

⁵¹ Corporate Duty of Vigilance Law 2017 (*Loi no 2017-399 du 27 mars 2017 relative au devoir de vigilance des sociétés mères et des entreprises donneuses d'ordre (1)*); Ordinance on Due Diligence and Transparency in relation to Minerals and Metals from Conflict-Affected Areas and Child Labour 2021 (*Verordnung vom 3. Dezember 2021 über Sorgfaltspflichten und Transparenz bezüglich Mineralien und Metallen aus Konfliktgebieten und Kinderarbeit (V/SoTr), AS 2021 847*).

⁵² Supply Chain Due Diligence Act 2021 (*Gesetz über die unternehmerischen Sorgfaltspflichten in Lieferketten, Bundesgesetzblatt Teil I, 2021, Nr. 46 vom 22.07.2021*).

⁵³ OHCHR, Guiding Principles on Business and Human Rights (2011) 13 <https://www.ohchr.org/sites/default/files/documents/publications/guidingprinciples_businesshr_en.pdf> accessed 14 September 2022; IBA, IBA Practical Guide on Business and Human Rights for Business Lawyers (2016) 17 <<https://www.ibanet.org/MediaHandler?id=d6306c84-c2f8-4c82-a86f-93940d6736c4>> accessed 14 September 2022.

and employment, and also require companies to undertake due diligence exercises, including with regard to their supply chain.⁵⁴ These have contributed to a ‘significant uptake’ in responsible business practices globally.⁵⁵ ESG-based investing is continuing to grow too, providing another impetus for companies seeking capital to become more involved in promoting good practices and policies across their business.⁵⁶ The nature of these frameworks is that they can indirectly affect and pass obligations onto even those companies that have not undertaken to respect them, where those companies demand suppliers also comply with responsible business practices.⁵⁷

G. The impact of socio-political and legislative developments on future class action tourism cases

Such developments — both legislative and non-legislative — could have substantial impacts on class action tourism cases in the coming years: (1) they could create new statutory bases for claims against parent companies in England that might replace tort claims, although the extent to which statutory heads of claim are used will be very case specific and will depend on the final drafting of the legislative proposals detailed above and the specific business models of potential defendants; (2) the generally more stringent legislation in European countries could mean fewer cases are brought in England, with claimants preferring to bring tort or statutory claims in those European jurisdictions. (3) Alternatively, they could provide a stronger basis for tort claims in England (for example, requirements — whether derived from statute or investor sentiment — to monitor, disclose, and implement policies related to human rights might aid claimants in showing parent companies were involved in managing their subsidiaries’ activities); (4) increased obligations may broaden the

⁵⁴ OECD, OECD Guidelines for Multinational Enterprises (2011) 19
<<https://www.oecd.org/daf/inv/mne/48004323.pdf>> accessed 14 September 2022.

⁵⁵ OECD, Stocktaking report on the OECD Guidelines for Multinational Enterprises (2022) 29

<<https://mneguidelines.oecd.org/stocktaking-exercise-on-the-oecd-guidelines-for-multinational-enterprises.html>> accessed 14 September 2022.

⁵⁶ Global Sustainable Investment Alliance, Global Sustainable Investment Review 2020 (2021) <<http://www.gsi-alliance.org/wp-content/uploads/2021/08/GSIR-20201.pdf>> accessed 14 September 2022.

⁵⁷ John Sherman, ‘Should a Parent Company Take a Hands-off Approach to the Human Rights Risks of its Subsidiaries?’ (2018) 19(1) Business Law International.

sort of business affected by claims beyond high-risk jurisdictions or industries;⁵⁸ or (5) fewer cases may be brought because companies implement more comprehensive governance policies that result in fewer instances of alleged wrongdoing.

What seems most likely is a combination of the above: greater due diligence requirements and the voluntary regimes that have been designed to achieve an improvement in corporate behaviour that reduces the number of claims brought. At the same time, where wrongdoing does occur and where related claims can be brought under new legislation, we should expect at least some claims that might otherwise have been brought under common law (as negligence claims, for example) being instead brought as statutory claims, whether in England or elsewhere. Finally, we might expect those tort claims that are brought to be more likely to succeed on the historically complicated issue of parental liability, given the increased obligations that large multinationals are likely to be under. Evidence of this third point can already be found. In *Vedanta*, Lord Briggs noted that ‘the parent may incur the relevant responsibility to third parties if, in published materials, it holds itself out as exercising that degree of supervision and control of its subsidiaries, even if it does not in fact do so’.⁵⁹ This is an important comment given some of the developments above are voluntary. Similarly, in *Okpabi*, group standards and policies were considered evidence of parental responsibility, including those produced for matters ‘subject to external stakeholder expectations and external disclosures’.⁶⁰ This is important as it means that even frameworks that have no enforceability mechanisms or are voluntary can be relevant in these cases. Similarly, in a 2016 case involving allegations that a mining company had failed to prevent police harming protesters at their mine, which was eventually struck out due to limitation issues, the claimants advanced their case on the basis that the defendant had fallen ‘below the standard of care set’ by the Voluntary Principles on Security and Human Rights to which it had signed up (though it was not clear that if they had fallen below these standards, this would amount to an

⁵⁸ James Brady, Marina Zarubin, Quinten De Keersmaecker, Clara Cibrario Assereto, ‘ESG Liability: Risks Increasing for Multinational Companies’ (2021) 32(5) PLC Magazine <<https://uk.practicallaw.thomsonreuters.com/w-031-0879>> accessed 5 August 2022.

⁵⁹ *Vedanta* (n 5) [53].

⁶⁰ *Okpabi* (n 3) [47].

automatic breach of a duty of care and expert evidence was not permitted on these principles).⁶¹

The issue of liability on the part of parent companies has been a vitally important one in class action tourism cases and the preceding analysis suggests that claimants may have an increased chance of success in demonstrating that liability. Regulatory changes and legislative proposals seem likely to force corporate groups to be more integrated and more likely to satisfy the requirements of case law, as recently articulated. However, when examining class action tourism cases more broadly, liability cannot be isolated, and the other crucial issue, in relation to which claimants may see less success in the future, is the question of jurisdiction.

II. JURISDICTION OF THE ENGLISH COURTS

Before turning to the current common law jurisdictional regime, it is worth considering the previous framework. This helps in understanding the rise of class action tourism and the different prospects for these cases now, and is potentially relevant for the future, as discussed by reference to the UK's application to join the Lugano Convention.⁶²

A. The Brussels Regime – the law

EU law has provided significant help to class action tourism claimants seeking to bring cases in England. The relevant jurisdictional regime immediately pre-Brexit was found in the regulation known as Brussels Recast.⁶³ This applied to proceedings instituted on or after 10th January 2015. In England, they continued to apply to cases brought before 11:00pm GMT on 31st December 2020.

⁶¹ *The Voluntary Principles on Security and Human Rights*

<<https://www.voluntaryprinciples.org/wp-content/uploads/2021/11/Voluntary-Principles-on-Security-and-Human-Rights-english-2.pdf>> accessed 18 December 2022; *Vilka v Xstrata Ltd* [2016] EWHC 2757 (QB), [2016] 11 WLUK 79 [10].

⁶² The Convention on Jurisdiction and the Recognition and Enforcement of Judgments in Civil and Commercial Matters [2007] OJ L339/3.

⁶³ Regulation (EU) 1215/2012 of the European Parliament and of the Council of 12 December 2012 on jurisdiction and the recognition and enforcement of judgments in civil and commercial matters (recast) [2012] OJ L351/1.

Crucially, for bringing class action tourism cases: (1) Article 4.1 of Brussels Recast states that ‘persons domiciled in a Member State shall, whatever their nationality, be sued in the courts of that Member State’; and (2) the judgment from the case of *Onusu v Jackson*⁶⁴ confirmed that contracting states could not ignore such provisions on the common law principle of *forum non conveniens*, the concept that another jurisdiction is a more appropriate forum for a claim.⁶⁵ Article 63.1 of Brussels Recast states that companies are domiciled in the place where they have their statutory seat, central administration, or principal place of business; English-incorporated companies would be domiciled in England under this article.

Article 4.1 (and earlier EU law) therefore allowed claimants to bring cases in England against English-domiciled companies and *Onusu* confirmed that courts could not depart from that rule.

As noted above, these cases generally involve parent and subsidiary defendants. Article 4.1 allowed the claim to be brought against the parent and this ‘anchor defendant’ allowed for the local subsidiary to be joined, as Article 6.1 of Brussels Recast states that jurisdiction for claims against non-domiciled defendants is decided by the relevant Member State’s laws (with some exceptions). In England, non-domiciled defendants can be served a claim form on the basis that they are a ‘necessary or proper party’ to the claim, this being one of the jurisdictional gateways under the Civil Procedure Rules (‘CPRs’).⁶⁶

B. Brussels Recast – challenging jurisdiction against a parent company

Despite this position, it was still necessary for there to be a case against the parent company; otherwise, the claim was liable to be stayed or struck out.

⁶⁴ Case C-281/02 [2005] QB 801. This involved a British claimant bringing a claim against a British defendant for an injury suffered in Jamaica; the courts ruled that the claimant was entitled to bring his claim in England, even though Jamaica seemed a more appropriate forum.

⁶⁵ While this case referenced earlier EU law, rather than Brussels Recast, the relevant articles are sufficiently similar that such case law was considered persuasive.

⁶⁶ Civil Procedure Rules, 6B PD 3.1.

Then, the lack of an anchor defendant would result in a lack of jurisdiction as regards the subsidiary.⁶⁷

Although it was relatively difficult to strike out a claim, this was still attempted, and the case of *Vedanta* is useful for illustrating the difficulties involved. In *Vedanta*, the defendants argued that there was no triable case against the anchor defendant, VR, but that, even if the court found there to be a triable issue against VR, such a claim was an abuse of EU law: it was being used merely as a tool to secure English jurisdiction for the case against KCM and would not have been brought otherwise.⁶⁸ These arguments were rejected; the courts suggested that if the VR claim had been brought solely for the purposes of bringing the claim against KCM, that may have constituted an abuse of EU law.⁶⁹ However, there was a real claim to be tried against VR and the claim had been brought at least in part for the genuine reason of obtaining damages; the court stressed that the scope to strike out claims as an abuse of EU law was narrow.⁷⁰

There are exceptions under EU law to Article 4.1, but these are generally irrelevant to the sort of class action tourism cases examined herein. One potentially relevant exception is where a related action is already pending in the court of a non-EU Member State.⁷¹ However, it has not always proved straightforward for defendants to use this exception: Article 34 gives courts discretion to stay proceedings, directing them to consider expediency, the administration of justice, and judgment enforceability.

The High Court would have used this exception to stay the claims brought against the English-incorporated BHP Group (UK) Ltd (known then as BHP Group Plc) which sought damages in relation to losses suffered when the Fundão Dam (operated by a joint venture part-owned by a BHP subsidiary) collapsed in Brazil in 2015. The High Court did not use this exception because they had already decided such claims were to be struck out because of abuse of process which, in large part, was also tied to the fact that parallel litigation was

⁶⁷ *AK Investment CJSC v Kyrgyz Mobil Tel Ltd* [2011] UKPC 7, [2012] 1 WLR 1804 [80], [82].

⁶⁸ *Lungowe v Vedanta Resources Plc* [2016] EWHC 975 (TCC), [2016] BCC 774 [74].

⁶⁹ *ibid*; *Vedanta* (n 5) [40].

⁷⁰ *Vedanta* (n 5) [24], [29].

⁷¹ Brussels Recast, art 34.

considered to be ongoing.⁷² However, the Court of Appeal overruled that judgment, finding that: (1) given the *forum non-conveniens* arguments of BHP Group (UK) Ltd's Australian-incorporated co-defendant failed (as discussed below), it would be inappropriate to stay the claims against BHP Group (UK) Ltd;⁷³ and (2) there was insufficient overlap between the collective proceedings that had been issued in Brazil and the proceedings in England for a stay to be ordered under Article 34.⁷⁴

Similarly, in a judgment relating to claims arising out of an oil spill in Nigeria, the High Court refused to grant a stay despite broadly similar proceedings in Nigeria for several reasons, though most importantly because there were slightly different issues to be determined in Nigeria and the Nigerian proceedings did not all involve the English-domiciled company.⁷⁵

In sum, after *Owusu*, EU law proved exceptionally helpful for claimants seeking to bring their claims in England. While there was still some scope to strike out or stay such claims, broadly, 'the debate [had] moved on', and it became very hard to argue that the court did not have jurisdiction to hear claims against a domiciled parent.⁷⁶

C. Brussels Recast – challenging jurisdiction against a subsidiary company

Attempts to strike out a claim against a non-domiciled subsidiary are, theoretically, easier to achieve, as CPR 6.37(3) provides that the court will not allow a claim form to be served out of the jurisdiction (notwithstanding the jurisdictional gateways) unless England is 'the proper place in which to bring the claim'. Frequently in class action tourism cases, the courts state that the foreign jurisdiction where the claimants are based and wrongdoing and losses occurred is ostensibly a more suitable forum than England, as discussed below. Notwithstanding such a finding, claimants have often relied on two factors that persuaded the court to hear such claims in England.

⁷² *Município De Mariana v BHP Group Plc* [2020] EWHC 2930 (TCC), [2020] 11 WLUK 91 [265].

⁷³ *Município De Mariana* (n 4) [298].

⁷⁴ *ibid* [282]-[284], [309]-[311].

⁷⁵ *Jalla v Royal Dutch Shell Plc* [2020] EWHC 459 (TCC), [2020] 3 WLUK 1 [245].

⁷⁶ Adrian Briggs, *Civil Jurisdiction and Judgments* (6th edn, Routledge 2015) para 2-304.

The first was that the claim against the parent company could continue in England and ‘to grant jurisdictional relief to some but not to others [i.e. to non-domiciled defendants but not to domiciled defendants] will fragment what ought to be conducted as a single trial’.⁷⁷ Previous case law viewed this as a trump card. Cooke J, in a 2003 case involving alleged breaches of a loan relating to the purchase of coffee in Kenya, stated that ‘the fact of continuing proceedings in England against other defendants on the same or closely allied issues virtually concludes the question’.⁷⁸ Leggatt J, in a 2013 case involving multiple English and Russian defendants who were connected to a Russian company which was alleged to have defaulted on its loans, similarly stated that, ‘[t]he real question is whether it is more appropriate to have that claim [against the Russian defendant] tried in Russia in addition to the claim against the other defendants in England raising all the same factual and legal issues ... the answer to that question is clearly no’.⁷⁹ However, in 2019, ‘after anxious consideration’ the Supreme Court in *Vedanta* overruled this interpretation, at least where the parent company was willing to accept the foreign state’s jurisdiction and irreconcilable judgments would only prejudice the claimants.⁸⁰ Instead, the right to claim against the parent in England and the desire to avoid conflicting judgments should be considered another relevant factor, not the decisive one.⁸¹

The second factor was whether justice could be obtained in the foreign jurisdiction. In *Vedanta* (as will be discussed in more detail below), the High Court found there was ‘cogent evidence that the claimants would not obtain access to justice in Zambia’ and the Supreme Court agreed (such that the Supreme Court’s different approach to conflicting judgments and jurisdictions described above was not determinative).⁸²

In sum, prior to 2021, if claimants could show an arguable case against the parent company, they could bring claims against that parent largely without fear of strike out on jurisdictional grounds. Jurisdictional challenges against the

⁷⁷ Lord Collins of Mapesbury and Jonathan Harris (eds), *Dicey, Morris and Collins on the Conflict of Laws* (15th edn, Sweet & Maxwell 2018) vol 1, para 12-033.

⁷⁸ *Credit Agricole Indosuez v Unicof Ltd* [2003] EWHC 2676 (Comm), [2004] 1 Lloyd’s Rep 196 [19].

⁷⁹ *OJSC VTB Bank v Parline Ltd* [2013] EWHC 3538 (Comm) [16].

⁸⁰ *Vedanta* (n 5) [75], [79].

⁸¹ *ibid* [84].

⁸² *Lungowe* (n 68) [169]; *Vedanta* (n 5) [98].

subsidiary are more complex. The foreign jurisdiction was often considered to be more appropriate, but the fact that the claim against the parent could continue in England was considered a strong argument against striking out the claim against the subsidiary. Before 2019, this was considered akin to a trump card although the strength of that argument was then weakened. A second crucial factor was whether the claimants could achieve access to justice abroad. The Brussels Regulations and EU case law have therefore been of substantial aid to class action tourism claimants, explaining, in part, the growth of these claims in recent years.

D. The Common Law

We turn now to the current regime and the prospects for future claims. In place of Brussels Recast, common law rules now apply to cases in England. These are examined first, before briefly considering what would happen if the UK were to join the Lugano Convention.

Under common law, effective service confers jurisdiction.⁸³ Service can be made on a registered UK company by, for example, serving at its registered address.⁸⁴ As before, a defendant properly served can be used as an anchor defendant for claiming against an overseas subsidiary. However, jurisdiction against that parent company can now be contested on *forum non conveniens* grounds. The two-stage test for this (the application of which is explored below) is found in *Spiliada Maritime Corp v Cansulex Ltd* (the ‘**Spiliada Test**’).⁸⁵ At the first stage, the defendant must show that another forum is clearly more appropriate than England; at the second stage, if stage one is met, a claim will be stayed unless the claimant can show (via ‘cogent evidence’) that justice requires that a stay not be granted.⁸⁶

⁸³ CPR 6.3 (n 66).

⁸⁴ Companies Act 2006, s 1139(1).

⁸⁵ *Spiliada Maritime Corp v Cansulex Ltd* [1986] UKHL 10, [1987] AC 460.

⁸⁶ *ibid* 478; there is some academic and judicial discussion about the line between these stages and whether certain factors – such as the practical availability of another forum that is theoretically available – fall into stage one or stage two; for a helpful summary of this discussion, which is not explored in detail here, see para. [334] onwards of *Municipio De Mariana*.

This test has been implicitly or explicitly applied in several recent class action tourism cases. For example, in *Vedanta*, the courts discussed these issues in relation to the claims against the subsidiary. On those issues that fall under stage one, the courts held that England was not the appropriate forum: Zambia was the location of the wrongdoing and losses, all the evidence and claimants were based there, and many did not speak English.⁸⁷

In the context of class action tourism and trans-national personal injury claims from the 1990s and early 2000s, stage two of the Spiliada Test, whether justice requires a stay not be granted, has often revolved around access to justice overseas.⁸⁸ In *Vedanta*, it was in relation to these issues that the claimants were successful. The courts held that justice did require that a stay should not be granted because of the risks to achieving justice in Zambia, namely: (1) the difficulty of funding these complex claims due to the claimants' poverty and the lack of funding mechanisms such as conditional fee arrangements, which are unlawful in Zambia; and (2) a lack of sufficiently capable legal teams.⁸⁹ In the High Court, Coulson J also noted evidence that KCM had been shielded from earlier litigation by political connections.⁹⁰

In order for a lack of funding to persuade a court against granting a stay, the absence of funding generally has to be such that the case could not be brought and justice not achieved; it is insufficient for there simply to be more funding in England. This can relate to the case being so complicated that it requires expert evidence or specialist legal advice.⁹¹ In contrast to *Vedanta*, in *Okpabi*, the High Court noted the presence of conditional fee arrangements and sufficient funding to pay for expert evidence in Nigeria, such that access to justice would not be denied by the case being heard in Nigeria rather than England.⁹²

⁸⁷ *Vedanta* (n 5) [85].

⁸⁸ Examples of these earlier cases include *Lubbe v Cape Plc* [2000] UKHL 41, [2000] 1 WLR 1545 and *Connelly v RTZ Corporation Plc* [1997] UKHL 30, [1997] 3 WLR 373.

⁸⁹ *Vedanta* (n 5) [89].

⁹⁰ *Lungowe* (n 68) [197].

⁹¹ *Connelly* (n 88) [31].

⁹² *Okpabi v Royal Dutch Shell Plc* [2017] EWHC 89 (TCC), [2017] 1 WLUK 495 [120]; such comments were obiter.

Further analysis of the Spiliada Test is provided in *Município De Mariana*. With regard to the first limb, the High Court had noted that Brazil was clearly the natural forum for claims against one of the two parent companies (which was not domiciled in England and so to which Article 4.1 did not apply). This was because: (1) the tort occurred in Brazil; (2) the governing law would be that of Brazil; (3) the Brazilian courts were practically more accessible for the majority of parties and witnesses, particularly given language considerations; and (4) the English courts could not rely on any pre-existing understanding of the case.⁹³ Regarding the second limb, the High Court was not persuaded that substantial justice could not be done in Brazil. They pointed out (among other objections) that redress may be available from other entities in Brazil, proceedings in England were unlikely to be quicker, and there was insufficient evidence that ‘impecuniosity would be a major factor in stifling legitimate claims’ in Brazil, where legal aid was available and a compensation scheme was in place. Further, the claimants would not have to pay a 30% success fee to their solicitors, as they would in England.⁹⁴

However, the Court of Appeal took a different approach. It stressed that, regarding the first stage of the Spiliada Test, it was not sufficient to simply refer to a jurisdiction as a whole as the more appropriate forum ‘when there are different forms of proceedings in different courts, each of which raise their own questions of appropriateness and suitability’.⁹⁵ In particular, the court distinguished between collective proceedings in Brazil and non-collective proceedings. This was important when considering the application of stage two of the Spiliada Test: the court held that (for various practical reasons relating to the diverse nature of the claimants and the process for bringing such proceedings) there was sufficient evidence to show a real risk that new collective proceedings were not an available forum for the claimants and that non-collective proceedings were inefficient, costly, and would result in a high risk of inconsistent judgments; the claimants therefore succeeded at the second stage.⁹⁶ Interestingly, the Court of Appeal took a provisional view that the claimants would have succeeded at stage one as well had they raised the argument that the collective proceedings process in Brazil was insufficient as it

⁹³ *Município De Mariana* (n 72) [238].

⁹⁴ *ibid* [255], [258].

⁹⁵ *Município De Mariana* (n 4) [343].

⁹⁶ *ibid* [352].

was asymmetric: it would ‘determine the rights of the parties if, but only if, it is favourable to the claimants’ (and might also take a very long time to resolve).⁹⁷

In addition, the Court of Appeal found the High Court judge had made a number of errors; in particular, the Court of Appeal held that the correct test was not ‘whether the claimants would not obtain substantial justice’, but rather whether there was a ‘real risk of them not doing so’, and also that redress from other parties in Brazil should not be taken into account.⁹⁸

Before concluding on the issue of jurisdiction and what a return to common law is likely to mean for those involved in class action tourism cases, it is important to briefly consider a possible, though increasingly unlikely, alternative regime, that of the Lugano Convention.

E. Lugano Convention

The Lugano Convention is similar to the pre-Brexit position described above: it replicates the legal regime prior to Brussels Recast and similarly states that parties should be sued in their place of domicile. Further, Article 1(1) of Protocol 2 states that Lugano Convention parties should ‘pay due account’ to relevant EU case law, which would include *Omnia*.⁹⁹

The UK was a member of the Lugano Convention as an EU member state and under the EU-UK Withdrawal Agreement during the transition period. It then ceased to be a member and, on 8th April 2020, applied to accede to the Lugano Convention in its own right. Iceland, Norway, and Switzerland have consented to the UK’s membership. However, the Commission announced on 4th May 2021 that it recommended the European Parliament and Council reject the UK’s application, arguing that it compromises the internal market and relations with European Free Trade Association countries.¹⁰⁰ In June 2021, the

⁹⁷ *ibid* [349]-[351]. The court noted at [350] that: ‘Whilst delay is normally a stage two factor, it can in our view be relevant to stage one when the relevant foreign proceedings may not conclude in any binding resolution of any of the parties’ rights and obligations.’

⁹⁸ *ibid* [355], [356].

⁹⁹ Lugano Convention, Article 1(1), Protocol 2.

¹⁰⁰ Commission, ‘Assessment on the application of the United Kingdom of Great Britain and Northern Ireland to accede to the 2007 Lugano Convention’ (Communication) COM (2021) 222 final.

Commission announced that the ‘European Union is not in a position to give its consent’ to the UK’s membership.¹⁰¹ Given unanimous consent of Lugano Convention parties is required and parties must ‘endeavour to give their consent’ within one year of any request, UK accession now seems unlikely.¹⁰² Further developments should be monitored, however, and it appears that the UK is still interested in acceding.¹⁰³

Ostensibly, acceding to the Lugano Convention would turn back the clock and much of the analysis above about the pre-Brexit regime would apply to jurisdiction questions in class action tourism cases. However, ‘pay due account’ is relatively weak wording and no penalties occur if EU case law is not followed: the Swiss Federal Tribunal has often departed from such case law, for example.¹⁰⁴ Such case law has been frequently criticised as inappropriate for common law jurisdictions and English judges, frustrated at being bound by *Onusu* in the judgments above, may also depart from it.¹⁰⁵ If the UK were to accede to the Lugano Convention, future case developments would need to be considered for indications of whether English courts appeared likely to follow EU case law or deviate from it on jurisdiction questions. It seems not unlikely that they would choose not to follow *Onusu* strictly, and therefore continue with the use of the Spiliada Test when determining jurisdictional challenges.

¹⁰¹ Commission, ‘Concerning the application of the United Kingdom of Great Britain and Northern Ireland to accede the 2007 Lugano Convention’ (Note Verbale), 22 June 2021.

¹⁰² Lugano Convention, art 72.

¹⁰³ Michael Cross, ‘UK ‘trying to re-engage’ on Lugano convention’ (2022) The Law Society Gazette

<<https://www.lawgazette.co.uk/practice/uk-trying-to-re-engage-on-lugano-convention/5114052.article>> accessed 18 December 2022.

¹⁰⁴ Burkhard Hess, ‘The Unsuitability of the Lugano Convention (2007) to Serve as a Bridge between the UK and the EU after Brexit’ (2018) 2 MPILux Research Paper Series 6

<https://www.mpi.lu/fileadmin/mpi/medien/research/MPEiPro/WPS2_2018_Hess_The_Unsuitability_of_the_Lugano_Convention_2007_to_Serve_as_a_Bridge_between_the_UK_and_the_EU_after_Brexit.pdf> accessed 13 September 2022.

¹⁰⁵ *ibid*; Collins (n 77), para 1-023.

F. The future of jurisdictional challenges

Assuming that the UK will not join the Lugano Convention in the immediate future (although, as discussed, even if it did, English courts might depart from EU case law and so the following analysis would also still be applicable), it seems likely that class action tourism cases involving jurisdictional challenges will now proceed as follows: (1) claimants serve on the English-domiciled parent company; (2) claimants seek to serve out on the subsidiary via the ‘necessary or proper party’ gateway; (3) claimants must demonstrate they have a sufficiently arguable case against the parent company which may be assisted by the case law and legislative developments referenced above; (4) defendants may contest jurisdiction against the parent, without which the subsidiary claim cannot be brought; (5) defendants are likely to satisfy the first limb of the Spiliada Test; and (6) claimants will have to rely on access to justice issues to succeed on the Spiliada Test’s second limb. The Supreme Court in *Vedanta* noted that the question of access to justice was, ‘prior to *Owusu v Jackson*, applicable in the context of an application for a stay of English proceedings against a defendant served within the jurisdiction’, and it seems it shall be again.¹⁰⁶

Given the importance likely to be placed on access to justice, it is worth examining this in a little more detail. In addition to the points mentioned above, from *Vedanta* and *Municipio De Mariana*, the courts have held that evidence of the following may be relevant: (1) a lack of ‘developed procedures’ for handling group claims, which may exacerbate funding issues;¹⁰⁷ (2) a political or ethnic dimension to the claim which could result in violence, intimidation, and abandonment of the case;¹⁰⁸ and (3) judicial corruption such that the claim might be abandoned (for example, if witness anonymity was breached resulting in intimidation).¹⁰⁹ In non-class action tourism cases, other issues with foreign

¹⁰⁶ *Vedanta* (n 5) [88].

¹⁰⁷ *Lubbe* (n 88) [30].

¹⁰⁸ *AAA* (n 9) [168], [169], albeit such comments were obiter.

¹⁰⁹ *ibid.*

courts have included: (1) excessive delay;¹¹⁰ (2) a time bar;¹¹¹ and (3) lack of impartiality more broadly.¹¹²

While claimants may seek to evidence one or other of these factors in their particular case, defendants may seek to persuade the courts that another jurisdiction is the appropriate forum by presenting evidence showing that there are no issues around access to justice or taking actions to mitigate those issues. Defendants may: (1) submit to that jurisdiction;¹¹³ (2) provide evidence from foreign lawyers that they will represent the claimants;¹¹⁴ (3) try to facilitate the judicial process overseas in some way (KCM's 'obdurate' nature was flagged in *Vedanta* as a reason why justice might be unachievable in Zambia, particularly given funding issues);¹¹⁵ and (4) admit some issues, particularly those requiring expert evidence. Again, the lack of admissions in *Vedanta* was considered relevant in a jurisdiction where funding was limited.¹¹⁶

Given that jurisdiction against a parent company can now be contested more easily than under the Brussels Recast regime and that that contestation has an impact on challenges against a subsidiary too, it seems possible that, under common law, more cases will be struck out as a result of jurisdictional challenges. When evaluating the merits of different jurisdictional tests in the context of strikeouts, various factors could be relevant including the efficiency of the process, the predictability of the outcome, and issues of comity (i.e. recognition of and respect for another jurisdiction's judicial system).¹¹⁷ But the most important issue is surely that which is so frequently the focus of the second limb of the Spiliada Test: whether claimants with legitimate grievances are able to pursue their claims and seek justice. Ostensibly, the fact that what may be otherwise legitimate and substantive claims are more likely to be struck

¹¹⁰ *The Jalakrishna* [1983] 12 WLUK 20.

¹¹¹ *Spiliada* (n 85) 483, 484.

¹¹² *Al-Koronky v Time-Life Entertainment Group Ltd* [2006] EWCA Civ 1123, [2007] 1 Costs LR 57.

¹¹³ *Radhakrishna Hospitality Service Pte Ltd v EIH Ltd* [1999] 2 Lloyd's Rep 249.

¹¹⁴ *Lungowe* (n 68) [187], although such evidence was not considered compelling.

¹¹⁵ *Vedanta* (n 5) [89].

¹¹⁶ *ibid* [94].

¹¹⁷ Ardavan Arzandeh, 'Should the Spiliada Test Be Revised?' (2014) *Journal of Private International Law* 10(1), 89 <<https://doi.org/10.5235/17441048.10.1.89>> accessed 14 September 2022.

out under the *Spiliada* Test than under the Brussels Recast regime might be a cause for concern.

However, the new status quo is potentially more nuanced than at first glance: claims are only likely to be struck out where they can properly be brought in a jurisdiction that is a more appropriate forum than England. As discussed above, appropriateness includes questions of practical accessibility, language barriers between claimants and the court, and the relevant law to be interpreted. The question the court asks is whether there is cogent evidence of a real risk of the claimants not obtaining substantial justice in another jurisdiction. It does not seem unreasonable to conclude that, where there is not such evidence, the case is better heard — and it is better that justice be seen to be done — in that other jurisdiction, rather than in England. This may be to the dissatisfaction of claimants who are making use of English lawyers and litigation funders and who will instead need to rely on potentially less experienced local lawyers and reduced funding. However, at least in principle, claimants should only be prohibited from bringing their claims in England where there are sufficiently capable lawyers and legal systems and sufficient sources of funding in their own country and the bringing of cases in local jurisdictions may help in developing local jurisprudence, experience, and frameworks for the successful resolution of these sorts of cases.

Further, in comparison to the previous regime (or the regime under the Lugano Convention were the courts to choose to follow EU case law), where cases that could have been heard in the claimants' own jurisdiction are instead required to be heard in England (due to Article 4.1 of Brussels Recast) and where claims against related parties could be brought in both an overseas jurisdiction and in England (following *Vedanta*) the common law test is potentially more efficient and less open to charges of what has sometimes been termed 'judicial chauvinism': the practice of considering the English courts better placed to hear a case that would, in fact, be better heard in another jurisdiction.¹¹⁸

¹¹⁸ *ibid* 102.

III. CONCLUSION

Class action tourism cases are very fact dependent. However, outlined above are principles that should assist practitioners. Two crucial issues arise in these cases: whether courts are likely to find liability on the part of the parent company, and whether they consider they have jurisdiction to hear the case.

Recent case law suggests English courts are unwilling to strike out claims on the issue of parental liability where claimants can present an arguable case and have given guidance on what evidence they find persuasive. It will be vital to: (1) analyse a parent company's structures and policies to see if case law indicates the courts will consider them capable of giving rise to liability; (2) follow relevant cases, particularly if they go to trial, to see how the courts' thinking develops; and (3) be aware of how market practice and legislative developments could affect claimants' cases. This last point may be particularly crucial: the corporate liability and accountability arena appears to be changing rapidly. It seems likely that this will advantage class action tourism claimants, but how exactly legislation, regulations, and the resulting behaviour of companies will affect this area of litigation remains to be seen.

On jurisdiction, access to justice overseas will likely be the crucial issue for determining whether English courts will hear these cases. Practitioners will need to understand: (1) the relevant foreign justice system more broadly; (2) how that justice system might influence their case specifically (for example, whether expensive medical expert evidence is required where funding is limited); and (3) what companies could do to persuade courts that justice can be done (for example, submitting to that foreign jurisdiction).

It seems possible that more class action tourism cases will fail on jurisdictional grounds than have in recent years or would have had the UK not left the EU; however, those that do not fail may be more likely to demonstrate a duty of care on the part of the defendant parent company, or successfully avoid strike out on the basis of potential liability. This does not ensure that such cases will achieve success at trial or via settlement. However, it probably does make such success more likely in comparison to the situation before the key Supreme Court cases mentioned above and, crucially, in comparison to a counterfactual where companies were not obliged or encouraged to engage with ESG

responsibilities in the way that they increasingly are. This field of litigation is fast-moving, complex, and requires an approach that looks at both important developments in case law, but also legislative and regulatory changes, all of which are sure to have an impact on this increasingly prominent trend in coming years.

Overcoming the dualism: how metaphorically-driven
conceptions of the public domain can be a barrier to
determining what intellectual property law is and ought
to be

Isabel Hahn*

ABSTRACT

It has been said that intellectual property and the public domain are 'paired together in a perpetual dance'. Yet while expressions like this portray the impression that intellectual property and the public domain are on an equal footing, it does not take long to discover that this is not the case. While intellectual property has been positively defined through centuries of legislation and case law, the public domain has witnessed no such equal treatment. Often lacking in positive definition, the public domain has also been poorly represented in the public debate. In an attempt to strengthen its position, scholars and judges have turned to the use of metaphors in order to capture and define the wealth of information and raw material from which we create, and to depict the function of the public domain in relation to intellectual property. However, the metaphors that have been used all share a common flaw: they obscure the fact that a literary device cannot adequately substitute for the lack of a clearly defined public domain in the first place. Therefore, to proactively safeguard the welfare of the public domain, we need to elevate the discussion beyond the limiting imagery and associations that metaphors offer and focus on the role that the public domain needs to play in a democratic society with a free market economy.

* LLB (LSE) '20. European Data Protection Supervisor (EDPS), Cabinet of the Supervisor. The views expressed by the author in this article are their own and do not represent the views of the EDPS.

INTRODUCTION

It has been said that intellectual property and the public domain are ‘paired together in a perpetual dance’.¹ Yet while expressions like this portray the impression that intellectual property and the public domain are on an equal footing, it does not take long to discover that this is not the case. While intellectual property has been positively defined through centuries of legislation and case law, the public domain has witnessed no such equal treatment. Often lacking in positive definition, the public domain has also been poorly represented in the public debate. In an attempt to strengthen its position, scholars and judges have turned to the use of metaphors in order to capture and define the wealth of information and raw material from which we create to depict the function of the public domain in relation to intellectual property. However, the metaphors that have been used all share a common flaw: they obscure the fact that a literary device cannot adequately substitute for the lack of a clearly defined public domain in the first place. Therefore, if we want to proactively safeguard the welfare of the public domain and the intellectual coherence of the law, we need to elevate the discussion beyond the limiting imagery and associations that metaphors offer. We must look beyond them to focus on the role that the public domain needs to play in a democratic society with a free market economy.

This article will begin by giving a quick overview of what intellectual property law is. The public domain is then introduced in order to show how its surrounding discourse is characterised by disagreement. Following this, some metaphors will be introduced that have been coined in an attempt to understand the role of the public domain and a brief critique of the metaphors that relate to natural resources, wastelands, and the life sciences will be offered. A substantial part of the article will then focus on the territorial metaphor of the public domain, which depicts the public domain as land. This metaphor will serve as a focal point, for it has been the most influential in shaping our conception of the public domain and has perpetuated the unfortunate view that intellectual

¹ Anupam Chander and Madhavi Sunder, ‘The Romance of the Public Domain’ (2004) 92 California Law Review 1331, 1339.

property should be the rule to which the public domain is the exception.² The conceptual misrepresentations and assumptions associated with the territorial metaphor will then be unravelled to show how they incorrectly depict the public domain, and how this depiction has been used by pro-commodificationists³ to further their agenda of expanding the intellectual property law regime. My argument follows in two parts.

In the first part, it is argued that the territorial metaphor poses a barrier to determining what intellectual property law is by oversimplifying and obscuring our understanding of the public domain. This will be shown by examining four misconceptions in our thinking on the public domain, triggered by the literal perception of a territory. The consequences of this metaphor for the way we conceptualise intellectual property law will then be discussed. First, the territorial metaphor triggers associations with the regulation of rivalrous and excludable goods. This is a barrier to determining what intellectual property law is because intellectual property law is our mechanism for assigning ownership and exclusive rights to non-rivalrous and non-excludable goods. Second, the territorial metaphor and our experiences with land prompt us to believe that the public domain has clear and tangible borders, which poses a barrier to the realisation that the barriers between intellectual property and the public domain are disputed, intangible, and constantly shifting. Third, the territorial metaphor depicts the public domain as static, which obscures the reality that intellectual property is dynamic and, most importantly, relational. Fourth, the territorial metaphor depicts a binary relationship between intellectual property rights and the public domain, which is a barrier to understanding that an intellectual creation of the human mind is capable of both being protected through intellectual property law and being part of the unprotected public domain simultaneously. This binary understanding of the public domain also legitimises the current distribution of intellectual property rights, masking the fact that this distribution inherently disadvantages indigenous and disempowered communities globally.

² Contrary to the position that ‘the public domain is the rule, copyright protection is the exception’ (COMMUNIA, ‘The Public Domain Manifesto’ (COMMUNIA) <<https://publicdomainmanifesto.org/>> accessed 10 February 2023).

³ A term coined by Cohen (Julie E Cohen, ‘Copyright , Commodification , and Culture: Locating the Public Domain’ from *The Future of Public Domain: Identifying the Commons in Information Law* (Kluwer Law International 2006) 121.).

In the second part of this article, it is argued that whether the territorial metaphor is a barrier to determining what intellectual property law *ought* to be is dependent on one's normative stance. The anti-commodificationist position is adopted, which seeks to limit the expansion of what is afforded intellectual property protection. It is submitted that here too the imagery and associations triggered by the territorial metaphor are a barrier to determining what the intellectual property regime ought to be. The territorial metaphor suggests that the intellectual property regime should continue to expand in order to safeguard the public domain. The case is made for why this should not be so.

The article concludes by bringing the discussion back to the multiple metaphors that have been used to describe the public domain in order to illustrate that while they might help to portray the public domain and intellectual property law in certain lights, the debate about what the public domain is will only continue as long as we continue to use metaphors. The suggestion is, therefore, made that the discussion be elevated beyond the use of metaphors to consider the central role that intellectual property law needs to play in promoting and safeguarding the sources of creative endeavour.

Throughout the article, reference will be made to a range of jurisdictions, including the United States, United Kingdom, and the European Union. This is done with the intention of depicting how different jurisdictions deal with the question of how the public domain should be defined. Moreover, it shows that the challenges of positively capturing what the public domain is and where its boundaries lie are common throughout.

I. A PRIMER ON INTELLECTUAL PROPERTY LAW

Intellectual property can be defined as 'the novel products of human intellectual endeavour'.⁴ Three principal branches of intellectual property will be discussed in this article: trademarks, copyright, and patents. Each branch differs with respect to what intellectual products it protects, why it assigns ownership to them, and under which circumstances it does so.

⁴Hector MacQueen and others, *Contemporary Intellectual Property, Law and Policy* (2nd edn, Oxford University Press 2010), 7.

Trademark law is concerned with signs capable of distinguishing goods or services⁵ and assigns ownership over a mark in order to ‘legally protect indications of commercial source’, letting a business profit from a ‘well-earned reputation for quality’.⁶ The exclusive right to a mark is provided when it can be shown that the mark has been used in the course of trade and that the public has come to recognize the mark as part of the producer’s signature.⁷ Copyright law is concerned with artistic or literary works⁸ and assigns ownership over a work in order to both incentivise the author to create more, and to compensate them for their creation.⁹ It does this when the author can demonstrate that the work is a product of their own intellectual creation. Finally, patent law is concerned with inventions¹⁰ and assigns ownership in order to incentivise the development and disclosure of inventions¹¹ (although this justification for patent law has been disputed).¹² The law grants a patent when an inventor can show that their invention is new, involves an inventive step (a formal requirement involving an improvement on the state of the art) and is capable of industrial application.

The regulation of these three branches form part of the umbrella term ‘intellectual property law’, which evolves in substance and scope over time. However, central to intellectual property is the notion that individuals do not just create out of thin air. The information that we draw on and the ideas that we have inevitably come from somewhere.

This general pool of knowledge, ideas, material and so forth, which are freely available to each and to all for use, constitute what is referred to as the

⁵ World Intellectual Property Organization, ‘Trademarks’ (*WIPO*) <<https://www.wipo.int/trademarks/en/>> accessed 10 February 2023.

⁶ Eric E Johnson, ‘Intellectual Property and the Incentive Fallacy’ (2012) 39 Florida State University Law Review, 626.

⁷ Jessica Litman, ‘The Public Domain’ (1990) 39 Emory Law Journal, 973.

⁸ World Intellectual Property Organization, ‘Copyright’ (*WIPO*) <<https://www.wipo.int/copyright/en/>> accessed 10 February 2023.

⁹ Litman (n 7) 970.

¹⁰ World Intellectual Property Organization, ‘Patents’ (*WIPO*) <<https://www.wipo.int/patents/en/>> accessed 10 February 2023.

¹¹ Edith Penrose and Fritz Machlup, ‘The Patent Controversy in the the Nineteenth Century’ (1950) 10 The Journal of Economic History 1, 10.

¹² With reference to the 4 arguments for the creation of patent rights, see Machlup & Penrose (n 11) 10.

‘public domain’. The public domain has been labelled as the ‘basis of our self-understanding’ and ‘the raw material from which new knowledge is derived and new cultural works created’.¹³ It is intertwined with intellectual property because ‘innovation captured as private property depends upon the existence of a rich public domain’.¹⁴ Intellectual property and the public domain, therefore, are ‘two sides of the same coin’¹⁵ with one inevitably affecting the other as its substance grows or shrinks.

II. DEFINING THE PUBLIC DOMAIN

Though the public domain is often conceived of as consisting of creative works to which no exclusive intellectual property rights apply, this definition is far from being absolute. Defining the public domain has been a task characterised by much disagreement across multiple dimensions. Some definitions are broader than others, viewing the public domain as ‘a sphere in which contents are free from intellectual property rights’¹⁶ or ‘a device that permits the rest of the system to work by leaving the raw material of authorship available for authors to use’.¹⁷ Some are narrower and define the public domain in the context of a specific intellectual property branch, such as Ginsburg’s definition of the public domain as ‘copyright’s constraining counterpart’.¹⁸ Temporally the definitions differ as well – some include all past and future inventions, some are more selective.

There is also disagreement about the substance of the public domain. Some view its substance in a ‘strict’ sense, where the public domain is comprised only of works whose intellectual property protection has expired and is thus property-free.¹⁹ Others view it in a ‘traditional’ sense, where it encompasses both works whose protections have expired and works that are not

¹³ COMMUNIA (n 2).

¹⁴ Chander and Sunder (n 1) 1340.

¹⁵ *ibid.*

¹⁶ Pamela Samuelson, ‘Mapping the Digital Public Domain: Threats and Opportunities’ (2003) 66 *Law and Contemporary Problems* 147, 149.

¹⁷ Litman (n 7) 968.

¹⁸ Jane Ginsburg, ‘Une Chose Publique: The Author’s Domain and the Public Domain in Early British, French and US Copyright Law’ (2006) 65 *Cambridge Law Journal* 636, 637.

¹⁹ Séverine Dussolier, ‘Scoping Study on Copyright and Related Rights and the Public Domain’ [2010] *World Intellectual Property Organization* 1, 7.

protected by intellectual property law.²⁰ Then there are those like Ginsburg, who in the context of copyright argue that this division between ‘strict’ and ‘traditional’ is arbitrary, as ‘even the public domain, in the strict sense of [expired] works, is not absolutely property-free’ as intellectual property rights can be invoked to protect the integrity of the work whose protection has expired.²¹

A third layer of disagreement emerges in the attempts to define the public domain’s function in relation to intellectual property. Ginsburg defines the function of the public domain as ‘enabling productive practices and exempting them from the exercise of an exclusive proprietary right’.²² Macmillan has also highlighted a similar relationship between the public domain and intellectual property rights, labelling the two as ‘symbiotic’²³ in the sense that they depend on one another to exist. While additional views depict the public domain as the antithesis of intellectual property,²⁴ Chander & Sunder and Lange have argued that we should keep away from this binary approach, claiming that understanding the public domain as the ‘opposite of property’ is misleading,²⁵ and the public domain is rather ‘a matter of public right, rather than simply the negative or obverse of intellectual property’.²⁶

These disagreements serve to emphasise the lack of an agreed upon definition of the public domain – both in depth and in breadth. Therefore, amidst this disagreement it is no wonder that we have turned to metaphors to help us capture our conception of the public domain. However, the metaphors we pick determine the ways in which we think about both the public domain and intellectual property law.

²⁰ *ibid* 6.

²¹ Ginsburg (n 18) 654.

²² *ibid* 9.

²³ Fiona Macmillan, ‘Many Analogies, Some Metaphors, Little Imagination: The Public Domain in Intellectual Space’ [2010] *Pòlemos: Potere e legittimà* 1, 9.

²⁴ James Boyle, *The Public Domain: Enclosing the Commons of the Mind* (Yale University Press 2008) 38.

²⁵ Chander and Sunder (n 1) 1343.

²⁶ Litman in reference to Lange (Litman (n 7) 132).

III. METAPHORS OF THE PUBLIC DOMAIN

One explanation for why we use metaphors is to understand and experience one kind of thing in terms of another.²⁷ They help us explain and understand a concept that is vague and unclear by substituting it with an image of something that is well-known and commonly experienced. Metaphors of the public domain have helped us construct a tangible understanding of it, which assists in illustrating where the dividing line lies between the proprietary and the public.²⁸ However, because of the inextricable link between the public domain and intellectual property, the way in which we conceptualise the public domain will inevitably impact our conception and decisions made with regard to intellectual property law.

This article will focus primarily on the territorial metaphor which portrays the public domain as a land. However, before this metaphor becomes the focus, some other metaphors that have been coined will be briefly acknowledged to explain how they can pose a barrier to determining what intellectual property law is and ought to be.

One set of metaphors depicts the public domain as a natural resource, such as water or air, in order to illustrate that resources in the public domain should remain public and available to all. These metaphors date back to nineteenth-century US intellectual property cases where common property in the public domain was expressly analogised with natural resources because it was believed that these types of resources were ‘incapable of private ownership’ and were owned by an ‘unorganised public at large’.²⁹ For instance, in *Holmes v Hurst*,³⁰ Justice Brown claimed that copyright did not entail the right to the use of certain words because words are ‘the common property of the human race and are as little susceptible of private appropriation as air or sunlight’.³¹ Similarly, Justice Brandeis claimed in his dissent in *International News Service v Associated*

²⁷ See George Lakoff and Mark Johnson, *Metaphors We Live By* (University of Chicago Press 1980) 4.

²⁸ Building on the definition given by Lakoff and Johnson on metaphors.

²⁹ Tyler Ochoa, ‘Origins and Meanings of the Public Domain’ (2002) 28 University of Dayton Law Review 215, 260.

³⁰ *Holmes v. Hurst*, 174 U.S. 82 (1899).

³¹ *ibid* [86].

*Press*³² that once the ‘noblest of human productions’ such as knowledge and ideas have been voluntarily communicated to others, they become ‘free as the air to common use’.³³

However, natural resources are depletable. And in Hardin’s view, the best way to protect a depletable resource is to privatise it in order to avoid a tragedy of the commons – a situation in which a public resource is depleted because of individuals acting in their own interest on that resource.³⁴ If we take the metaphor of the public domain as a natural resource too literally, we may fall prey to the assumption that the public domain is depletable and should be extensively privatised in order to safeguard it. In other words, the metaphor of a natural resource propagates the belief that the only way to ‘protect’ the ‘depletable’ nature of the public domain is through assigning property rights to individuals. However, this in itself limits our understanding of the nature of intellectual property law, insofar as it justifies the indefinite expansion of the intellectual property regime in order to safeguard the public domain. The danger with this is that it would lead to what Heller & Eisenberg deem a ‘tragedy of the anticommons’ scenario, where privatisation leads to the underuse of resources in the public domain through too many fragmented and overlapping intellectual property rights, stifling any kind of innovation and creation.³⁵

However, what the tragedy of the commons and anticommons scenarios have in common is that they both lead to the trap of associating the public domain with a natural resource. They both perpetuate a narrative that projects the characteristics of the metaphor onto our working definitions. This is problematic insofar as it constrains our thinking and understanding of the public domain to the parameters of the metaphor itself, which does not capture the reasons why we choose to assign ownership to intellectual property goods.

³² *International News Service v. Associated Press*, 248 U.S. 215 (1918).

³³ *ibid* per Justice Brandeis at 248 U.S. 250.

³⁴ Garrett Hardin, ‘The Tragedy of the Commons’ (1968) 162 (3859) *Science*, 1245.

³⁵ Michael A Heller and Rebecca S Eisenberg, ‘Can Patents Deter Innovation? The Anticommons in Biomedical Research’ (1998) 280(5364) *Science* 699.

A second set of metaphors view the public domain simply as a 'wasteland'.³⁶ This camp of metaphors includes expressions like the public domain as a 'fallow land'³⁷ or as 'gummy residue',³⁸ and normally portray the public domain as comprising only of works whose intellectual property protection has expired. However, an obvious shortcoming of these metaphors is that they fail to account for how the term 'public domain' has come to refer to all things in the public domain, including creations which have been deemed as not original enough to be awarded intellectual property protection, as opposed to only creations who have had their protections expire. Furthermore, the negative connotations associated with the metaphors 'fallow land' or 'wasteland' suggest that what is in the public domain is no longer useful. Yet this is incorrect, because our fundamental building blocks of knowledge and creation rest in the public domain, such as the English language or general plots and facts.³⁹ Thus, these metaphors constitute a barrier to determining what intellectual property law is as they implicitly suggest that everything 'valuable' and 'useful' is privatised, while in fact the things that are most valuable and useful to us all rest in the public domain.

A third set of metaphorically-driven conceptions of the public domain is constructed around the life sciences, and seeks to replace our thinking of the public domain with notions such as 'information ecology'⁴⁰ or 'cultural software'⁴¹ which seek to explain creative processes in a way that embraces complexity and interdependence. Though they better account for how information flows affect one another, Cohen points out that a common shortcoming to them is that they 'betray a worrisome tendency towards animism'⁴² in the sense that they elevate the role of information to being separate from human agency.

³⁶ Séverine Dusollier, 'The Public Domain in Intellectual Property: Beyond the Metaphor of a Domain' in PL Jayanthi Reddy (ed.), *Intellectual Property and Public Domain* (ICFAI University Press 2009) 3.

³⁷ *ibid.*

³⁸ Boyle (n 24) 40.

³⁹ *ibid.* 38.

⁴⁰ James Boyle, 'A Politics of Intellectual Property: Environmentalism for the Net?' (1997) 47 *Duke Law Journal* 87, 93.

⁴¹ JM Balkin, *Cultural Software: A Theory of Ideology* (Yale University Press 1998) Chapter 1.

⁴² Cohen (n 3) 142.

Although there are more ways to metaphorically describe the public domain, such as Cohen's understanding of a 'cultural landscape'⁴³ where the public domain is seen as a set of cultural and creative practices, attention is now shifted to the territorial metaphor.

IV. THE TERRITORIAL METAPHOR: PUBLIC DOMAIN AS LAND

4.1 Origins of the territorial metaphor

Expressions seeking to portray the public domain as territorial date back to early US intellectual property cases, in particular the decisions of *Webster's Dictionary*⁴⁴ and *Singer Manufacturing Co. v. June Manufacturing Co.*⁴⁵ In *Webster's Dictionary*, Supreme Court Justice Samuel Miller declared that once copyright had expired, works 'shall go to the public and become public property'.⁴⁶ Subsequently in the *Singer* case, the Supreme Court quoted Justice Miller's discussion of 'public property' to conclude that the name 'Singer', along with the public ownership of its sewing machines, had now fallen 'into the domain of things public'.⁴⁷ The term 'public domain' was then codified in the US 1909 Copyright Act, where Section 7 expressly excluded protection for 'works in the public domain'.⁴⁸

Moreover, the term 'public domain' was initially used in relation to the disposition of publicly owned lands, which were geographically separate places not subject to direct private appropriation. However, despite Ochoa pointing out that the phrase 'public domain' evokes the notion of public lands and in doing so 'mischaracterizes the public domain in intellectual property',⁴⁹ modern references to the public domain as land have persisted: Cohen has addressed the public domain as a 'geographically separate preserve',⁵⁰ and further contemporary scholarship has also referred to it as 'a landscape where each

⁴³ Cohen (n 3) 157.

⁴⁴ *Merriam v Holloway Publishing Co.*, 43 F. 450 (1890).

⁴⁵ *Singer Manufacturing Co. v. June Manufacturing Co.*, 163 U.S. 169 (1896).

⁴⁶ Cohen (n 3) 126.

⁴⁷ *Singer* (n 45) [185].

⁴⁸ Cohen (n 3) 126.

⁴⁹ Ochoa (n 29) 267.

⁵⁰ Cohen (n 3) 124.

person can reap the riches found in the common'.⁵¹

These early judicial, legislative and academic developments are significant because they depict how the public domain was seen as territorially separate and in opposition to intellectual property from early on.

However, as the following discussion will show, there are several distinguishing characteristics between land and the public domain that pose a barrier to determining what intellectual property law is and ought to be.

4.2 The territorial metaphor triggers an association with producing and safeguarding rivalrous and excludable goods

When one thinks of land, one thinks of the rivalrous and excludable goods it yields. In other words, one thinks of goods that are finite and limited. Thinking about the public domain through the lens of the land metaphor brings with it the experience of working hard to produce a finite number of goods that are grown from a limited number of resources. Yet, we have come to mistakenly equate the mechanism of producing and safeguarding tangible goods with the mechanism of producing and safeguarding intangible goods.

The territorial metaphor masks the crucial distinction between land as giving rise to rivalrous and excludable goods, and the public domain as giving rise to non-rivalrous and non-excludable goods. Consequently, it has also tempted us into mistakenly applying the logic behind regulating tangible goods to the logic for regulating intellectual goods.⁵²

Labour theory of property,⁵³ for instance, holds that a tangible property right is warranted because people have a natural right to the fruits of their labour. Moreover, it justifies the grant of such a property right in rivalrous goods 'in order to prevent the tragedy of the commons'.⁵⁴ Assigning people ownership over their tangible goods both stimulates them to produce more and

⁵¹ Chander and Sunder (n 1) 1332.

⁵² Brian L Frye, 'IP as Metaphor' (2015) 18(3) *Chapman Law Review* 735, 757.

⁵³ As intended by John Locke in *Second Treatise of Government* (edited by C. B. Macpherson, Hackett Publishing 1980) Section 6.

⁵⁴ Frye (n 52) 741.

simultaneously ensures that their resources will not become depleted by virtue of being exploited in common by all. Indeed, this was the logic behind the enclosure of the commons which took place in fifteenth century England, where it was thought that in order to most efficiently utilise the common land and stimulate production it must be privatised and converted into private property.⁵⁵

However, such a labour theory justification for property rights does not hold in the context of non-rivalrous goods, because there *is* no such tragedy of the commons in relation to intellectual goods.⁵⁶ While tangible goods might be depleted if they are exploited in common, conversely, intangible goods like ideas only *multiply* in quantity when they are exploited in common. Therefore, the idea here that the exploitation of intangible goods may lead to a tragedy of the commons scenario is in itself fallacious. Yet despite this, intellectual property law still ‘creates artificial scarcity so that the holder of the right can generate returns from’ their creation.⁵⁷ One can only question why this has been done, presumably in order to create an additional incentive to produce more intellectual fruits of labour, falsely using the economic rationality of labour theory as a justification.

Furthermore, we have transplanted the natural rights claim to ownership from the context of tangible goods onto our justification for assigning intellectual property rights. An example of this is the romantic version of authorship, which holds that an author has a claim to ownership and control over their work because of the ‘individualised act through which the personal genius of the author [was] brought to the world’.⁵⁸ This not only assumes that authors are *entitled* to rights for their creation, but also that in order to compensate them for their creation we should establish property rights in their

⁵⁵ Donald N McCloskey, ‘The Enclosure of Open Fields: Preface to a Study of Its Impact on the Efficiency of English Agriculture in the Eighteenth Century’ (2010) 32 *The Journal of Economic History* 15, 19.

⁵⁶ *ibid.*

⁵⁷ Siva Thambisetty, ‘It’s not just about patents on COVID vaccines: why I am not celebrating World Intellectual Property Day’ (*LSE*, 26 April 2021) <<https://blogs.lse.ac.uk/covid19/2021/04/26/its-not-just-about-patents-on-covid-vaccines-why-i-am-not-celebrating-world-intellectual-property-day/>> accessed 10 February 2023.

⁵⁸ Carys Craig and Ian Kerr, ‘The Death of the AI Author’ (2021) 52(1) *Ottawa Law Review* 31, 77.

work. But critics of this have been quick to object to assigning tangible property rights to intangible goods. Jefferson, for instance, thought that the ‘fugitive fermentation of an individual brain’ could not be the subject of a natural right to exclusive and stable property.⁵⁹ Similarly, Condorcet believed that property rights in a book ‘could not be compared to those in a field or piece of furniture which could only be occupied or used by one person at a time’.⁶⁰ Yet despite this, the natural rights approach to intellectual property is still very much alive: the investment function of the trademark is ‘framed around the idea that since the trademark owner invested in the mark ... [they] ought to reap the rewards of this investment and should be allowed to prevent others from misappropriating the value which results’.⁶¹

The continued use of the territorial metaphor only further perpetuates this natural rights/labour justification for intellectual property rights. This has been exploited by those looking to expand the intellectual property regime. Unfortunately, the other prominent justification for intellectual property rights — the economic justification theory, otherwise known as incentive theory — is also arguably flawed.⁶² The consequence of this is that it leaves those who are looking to safeguard the public domain with almost no ground to stand on, as they are unable to point to a justification that advocates for the limited granting of intellectual property rights.

The economic justification for intellectual property holds that because artistic works and technological innovations are non-rivalrous and non-excludable goods, they are vulnerable to free riding.⁶³ As a consequence, people will be disincentivised from engaging and investing in intellectual labour and will consume the innovations created by others, leading to market failure. Thus, intellectual property rights provide the external rewards needed to

⁵⁹ Thomas Jefferson, ‘The Founders’ Constitution, Article 1, Section 8, Clause 8. Thomas Jefferson to Isaac McPherson’ <https://press-pubs.uchicago.edu/founders/documents/a1_8_8s12.html> accessed 10 February 2023.

⁶⁰ Boyle (n 24) 31.

⁶¹ Luke McDonagh, ‘From Brand Performance to Consumer Performativity: Assessing European Trade Mark Law After the Rise of Anthropological Marketing’ (2015) 42 *Journal of Law and Society* 611, 618.

⁶² Lionel Bently and others, *Intellectual Property Law* (5th edn, Oxford University Press 2018) 5.

⁶³ *ibid.*

incentivise people to create and invest in intellectual goods, which remedy the free rider problem and reduce transaction costs.⁶⁴ Under this theory, intellectual property is a temporary, state-created monopoly given to encourage innovation⁶⁵ and is justified only to the extent that ‘each grant of rights provides the necessary incentive to cause its recipient to provide a social benefit and no more, and only if the social benefit is larger than the social cost of the right’.⁶⁶

Intellectual property’s need to encourage invention and artistic production has been referred to as its ‘*raison d’être*’.⁶⁷ However, this incentive theory is coming under fire as new behavioural economics research reveals that not only is incentive not necessary to encourage individuals to create, but that ‘external rewards can also actually *disincentivize* creative labourers’.⁶⁸ The digital revolution, for instance, has led to an increase in the means available to individuals to create. An example of such a means is ‘commons-based peer production’, which is a model of social innovation based on collaborative production.⁶⁹ Significantly commons-based peer production confirms that individuals will create without any external incentive or expectation of compensation. Patent law, for instance, is too often justified on the grounds of incentive theory, yet empirical data is revealing that ‘many patented inventions would have been developed even in the absence of the patent system’.⁷⁰ Cohen has also criticised the application of incentive theory to copyright law, arguing that economic modelling and markets insufficiently understand creativity, as ‘creativity is a social phenomenon that is both broader than and antecedent to the market exchange of goods and services’.⁷¹

Returning to the discussion of the territorial metaphor, our current economic approach to intellectual property law reveals the conviction that propertization incentivises productivity. It is therefore no wonder that we find ourselves on the path to what Boyle refers to as ‘the second enclosure

⁶⁴ Frye (n 52) 737.

⁶⁵ Boyle (n 24) 20.

⁶⁶ Frye (n 52) 747.

⁶⁷ Johnson (n 6) 635.

⁶⁸ *ibid* 643.

⁶⁹ Yochai Benkler and Helen Nissenbaum, ‘Commons-Based Peer Production and Virtue’ (2006) 14 *Journal of Political Philosophy* 394, 394.

⁷⁰ Johnson (n 6) 665.

⁷¹ Cohen (n 3) 137.

movement',⁷² privatising that which was previously part of the public domain under the false pretence that this best encourages creativity and prevents the public domain from depletion. However, this is a mistaken view, because an increased use of the public domain will not lead to less creativity, as the economic approach would suggest. Instead, it could be that individuals will increasingly find inspiration to create ideas from that which already exists. This view that we ought to privatise the public domain to best protect it obscures the reality that the public domain is expanded by the sharing, combining, and paraphrasing of ideas.

However, because the economic narrative that propertization incentivises productivity is so strong, it is difficult to garner support for the opposing view: that not affording intellectual property rights to that which is in the public domain can actually lead to more creation. This idea will be explored in the section that deals with the question of what intellectual property law ought to be; however, for now, focus will be shifted to the second characteristic associated with the metaphor of land.

4.3 The territorial metaphor triggers the image of clear and tangible boundaries

A second conceptual characteristic associated with a territory is that it has clear and tangible boundaries, allowing us to distinguish between what is part of it and what is not. However, this image of the public domain masks the fact that there is no clear and tangible border around the public domain that can be invoked to protect it. Because intellectual property is intangible, the law must supply alternative concepts that can assume the role of physical boundaries to divide it from the public domain. Yet, as will be demonstrated with reference to copyright law, there are no such clearly defined boundaries, and the current boundaries are poorly defined and continuously shifting. One consequence of this is that pro-commodificationists have been able to assert the boundaries of the law broadly in order to further their agenda that the law ought to continue to expand the intellectual property regime.

In European copyright law, what was relied upon in the place of physical borders (prior to the case of *Infopaq International A/S v Danske Dagblades*

⁷² Boyle (n 24).

*Forening*⁷³) was the concept of originality.⁷⁴ Yet, as Litman has pointed out, the threshold of originality was ‘inherently unascertainable’⁷⁵ and was a poor substitute for tangible boundaries. Since authors ‘necessarily reshape the prior works of others, a vision of authorship as original creation from nothing ... is both flawed and misleading’,⁷⁶ says Litman. However, the law has since moved on from the standard of originality and has now adopted a new threshold for protected works, which is the ‘author’s own intellectual creation’.⁷⁷

This has been part of what Griffiths labelled a shift towards a ‘dematerialised model’ of copyright, in which the law looks at the ‘immaterial, malleable essence’ of a work in order to determine the scope and attribution of rights instead of the boundaries of the material form within which the work was first recorded.⁷⁸ While this ‘immaterial essence’ used to be measured by the thresholds of ‘labour and skill’,⁷⁹ ‘expression’,⁸⁰ and ‘originality’,⁸¹ *Infopaq* has shifted this standard. *Infopaq* concerned an action for copyright infringement of an eleven word extract of a newspaper. The aspect of the judgment which drew the most attention was the Court’s decision that the ‘author’s own intellectual creation’ (or the ‘creativity’ standard) applied to all forms of copyright work protected under the Information Society Directive, rather than only works such as computer programmes, databases or photographs.⁸²

The fact that the Court never defined what an ‘author’s own intellectual creation’ is further blurs this distinction between what is to be afforded copyright protection and what is not.⁸³ This has the potential to pose a serious threat to the safeguarding of the public domain, especially because following *Infopaq* the standard for infringement has also been lowered.

⁷³ Case C-5/08 *Infopaq International A/S v Danske Dagblades Forening* [2009] ECR I-6569.

⁷⁴ Litman (n 7) 1000.

⁷⁵ *ibid* 969.

⁷⁶ *ibid*.

⁷⁷ *Infopaq* (n 73).

⁷⁸ Jonathan Griffiths, ‘Dematerialization, Pragmatism and the European Copyright Revolution’ (2013) 33 *Oxford Journal of Legal Studies* 767, 767-768.

⁷⁹ *Sawkins v Hyperion Records Ltd* [2005] EWCv Civ 565 [85].

⁸⁰ *Designers Guild Ltd v Russell Williams (Textiles) Ltd* [2000] 1 WLR 2416 (HL) 2431.

⁸¹ Copyright Design and Patents Act 1988, s 21; *Sillitoe v McGraw-Hill Book Co (UK) Ltd* [1983] FSR 545 (Ch).

⁸² *Infopaq* (n 73) [35].

⁸³ Griffiths (n 78) 789.

For instance, in the United Kingdom, under the CDPA 1988,⁸⁴ infringement takes place whenever an act is committed in relation to the whole or a 'substantial part' of the work that is restricted by the copyright.⁸⁵ No liability used to arise for the reproduction of an 'insubstantial' part of the work. However, following *Infopaq*, the standard for infringement had become stricter, and infringement arises whenever 'elements which are the expression of the intellectual creation of the author of the work' are reproduced.⁸⁶ This seems to leave 'no apparent scope for a defendant to reproduce any of a work's 'creativity' without infringing copyright'.⁸⁷ McDonagh has commented that this has been particularly burdensome for the music industry, as now it is 'more difficult than ever to determine when a case of musical infringement has occurred', since it seems that 'even a short progression of chords could be protected' by copyright if it is deemed to be part of the intellectual creation of the author.⁸⁸

Whether this blurred border between copyright and the public domain is desirable depends on whose interests are under consideration. For pro-commodificationists, it is extremely advantageous that the *Infopaq* decision keeps the romantic notion of the author alive by suggesting that authors should have a claim to ownership over anything that is their own creation, as it can be used to expand the regime of copyrighted works.

Trademark and patent law too have witnessed a similar blurring of the boundary between what is protected by intellectual property law and what is not. This consequently has also made it harder to positively assert a division between the private and the public in order to protect the public domain. For instance, in trademark law the boundary for what could be protected as a trademark used to be set at signs that were capable of being visually represented. This graphic representation requirement⁸⁹ was a way to resolve 'the apparently

⁸⁴ Copyright Design and Patents Act 1988.

⁸⁵ *ibid* s16(3)(a).

⁸⁶ *Infopaq* (n 73) [39].

⁸⁷ Griffiths (n 78) 787.

⁸⁸ Luke McDonagh, 'Is the Creative Use of Musical Works without a Licence Acceptable under Copyright Law?' (2012) 4 *International Review of Intellectual Property and Competition Law* 401, 419.

⁸⁹ First Council Directive 89/104/EEC to approximate the laws of the Member States relating to trade marks [1988] OJ L40/1.

intractable problem of identifying and demarcating the boundaries of intellectual property law'.⁹⁰ However, *Sieckmann v Deutsches Patent*⁹¹ made it clear that it is no longer necessary for a sign to be capable of graphic representation in itself in order to constitute a trademark. In other words, a non-visual sign can now be a trademark too.

Sieckmann indicated that a sign not capable of visual perception might still meet the standard of graphic representation as long as it could be represented in a way that was 'clear, precise, self-contained, easily accessible, intelligible, durable and objective'.⁹² This latter list of requirements, known as the *Sieckmann* criteria, was then used in *Libertel v Benelux*⁹³ to conclude that a single colour mark could also be a trademark. It was also used by Advocate General Ruiz-Jarabo Colomer in *Shield Mark v Joost Kist*⁹⁴ to determine that a series of sounds could be trademarks if they could be represented by musical notation. This has blurred the boundaries of the law because it can no longer be said 'that representation alone determines the precise subject of protection'.⁹⁵

Similarly, the boundary between invention and discovery has been 'blurred to a dangerous point'⁹⁶ in patent law. For instance, Article 3(2) of the European Directive on Biotechnology states that 'biological material which is isolated from its natural environment or produced by means of a technical process may be the subject of an invention even if it previously occurred in nature'.⁹⁷ This isolation test blurs the border between invention and discovery because it is questionable whether 'the act of isolation and characterization of a naturally occurring substance is really that different from the mere finding of the substance'.⁹⁸

Thus, the metaphor of the public domain as land as currently

⁹⁰ Robert Burrell and Michael Handler, 'Making Sense of Trademark Law' (2003) 4 Intellectual Property Quarterly 388, 388.

⁹¹ Case C-273/00 *Sieckmann v Deutsches Patent* [2002].

⁹² *ibid* [55].

⁹³ Case C-104/01 *Libertel Groep BV v Benelux-Merkenbureau* [2003].

⁹⁴ Case C-283/01 *Shield Mark BV v Joost Kist* [2003].

⁹⁵ Burrell and Handler (n 90), 408.

⁹⁶ Dusollier (n 36) 11.

⁹⁷ Council Directive 98/44/EC of the European Parliament and of the Council of 6 July 1998 on the legal protection of biotechnological inventions [1998] OJ L 213/13.

⁹⁸ Bently and others (n 62) 495.

constructed obstructs the understanding that, in reality, there is no clear and tangible border around the public domain that can be invoked to protect it. The consequence of the lack of a positively asserted border is that it is hard to fight back against the encroaching intellectual property regime.

4.4 The territorial metaphor elicits the image of a static construct

The third characteristic associated with a territory is its static nature. However, this is deceptive insofar as it falsely suggests that the public domain and intellectual property law are static too. Unlike territory or land, the public domain is not fixed; rather, it is dynamic and relational, growing and shrinking in size at different times and in different countries. Sometimes the public domain grows, for instance when copyright or patented works expire, or following decisions like *Feist Publications, Inc. v. Rural Telephone Service*⁹⁹ in the United States, where it is held that uncreative compilations of facts cannot be protected by US copyright law. At other times the public domain shrinks, such as when the EU enacted its EU Database Directive¹⁰⁰ which required that Member States protect the contents of databases, or when the US Courts ruled that business methods could in fact be patented.¹⁰¹ However, the consequence of adopting a static interpretation of the public domain is that one can remain blissfully ignorant of the fact that whenever the intellectual property regime grows, the public domain shrinks. The static understanding of the public domain thus poses a barrier to determining that intellectual property law should expand only when necessary, in order to prevent the public domain from unduly shrinking.

An example of how the intellectual property regime has grown at the cost of the public domain is the growth of trademark law to protect brand dimension associated with a successful trademark.¹⁰² This happened in the case *L'Oréal v Bellure*,¹⁰³ where the Court of Justice of the European Union declared that the functions of a trademark included not only its origin function, 'but also

⁹⁹ *Feist Publications, Inc. v. Rural Telephone Service Co.*, 499 U.S. 340 (1991) 363-64.

¹⁰⁰ Council Directive 96/9 of 11 March 1996 on the Legal Protection of Databases [1996] OJ L77.

¹⁰¹ *State Street Bank & Trust Co. v. Signature Financial Group, Inc.*, 149 F.3d (Fed. Cir. 1998) 1368, 1375.

¹⁰² Dev S Gangjee, 'Property in Brands' (2013) 8 LSE Legal Studies Working Paper 2.

¹⁰³ Case C-487/07 *L'Oréal SA v Bellure NV* [2009] ECR I-5185.

its other functions, in particular that of guaranteeing the quality of the goods or services in question and those of communication, investment or advertising'.¹⁰⁴ These additional functions are now also capable of being protected even where the origin function of the trademark remains unaffected.¹⁰⁵ This is problematic because it has been suggested that these functions are so prominent largely due to the consumer's engagement with the mark;¹⁰⁶ indeed, 'consumers who participate and amplify different meanings contribute enormously to the success of a brand'.¹⁰⁷

Previously, the brand value and image generated by consumer engagement with a mark would have rested in the public domain. Now, the brand owner enjoys all of the rights associated with the consumer's labour. Not only is this appropriating hidden labour, but it also goes to show how unpredictable the growth of trademark law is. The consequence of this is that the public domain has shrunk again simply because consumers effectively engaged with a brand.

4.5 The territorial metaphor depicts a binary relationship between intellectual property right and the public domain

The public domain is also not in binary opposition to intellectual property as the territorial metaphor would suggest. As academics like Samuelson have pointed out, there are intellectual creations 'that have been treated as part of the public domain for some but not all purposes'.¹⁰⁸ The ability of a private resource to be part of the function of the public domain, because it can be freely used in some circumstances, is reflective of what Dussolier calls a 'functional public domain', where the availability of the material in the public domain depends on the circumstance.¹⁰⁹ This can be best illustrated with reference to how copyright and patent law allow free use of a protected work or invention in certain situations.

¹⁰⁴ *ibid* [58].

¹⁰⁵ McDonagh (n 61) 612.

¹⁰⁶ Gangjee (n 102) 8.

¹⁰⁷ Thambisetty (n 57).

¹⁰⁸ Samuelson (n 16) 149.

¹⁰⁹ Dussolier (n 36) 7.

For instance, the exceptions in copyright law are an example of how the public domain is embedded in the proprietary regime itself. There are a number of fair dealing purposes under the CDPA 1988, which include the right to use a copyrighted work for caricature, parody or pastiche purposes¹¹⁰ and for criticism or review purposes.¹¹¹ In *Pro Sieben Media v Carlton Television*,¹¹² for instance, Carlton Television was able to claim fair dealing in order to use a 30-second extract from an interview broadcasted on ProSieben Media for the purposes of criticising chequebook journalism. This is an example of how a copyrighted work may be ‘freely’ used despite its proprietary status, depicting how the intellectual property regime is also dynamic rather than static.

In patent law too, there are exceptions that allow the use of a patented invention in certain instances. In the UK, Section 60(5)(a) and 60(5)(b) of the UK Patents Act 1977 (as amended), state that acts that are done for private and non-commercial purposes or experimental purposes will not constitute infringement of a patent. This is another example of how an intellectual property right is relational depending on the instance in which it is being used. This can be further illustrated by reference to Article 31 of the Agreement on Trade-Related Aspects of Intellectual Property Rights (1994), under which a government can issue a compulsory licence which gives producers the right to manufacture a patented product or process without the consent of the patent holder.

Indeed, compulsory licences exemplify how patented products may become temporarily part of the public domain, thus attesting to the fact that the intellectual property regime is not binary. For example, in March 2020 the Israeli government issued a compulsory licence for the HIV drug Kaletra so that it could be manufactured more widely for use in COVID-19 treatment trials.¹¹³ This was done after the Israeli Ministry of Justice announced that the pharmaceutical company AbbVie, which held the patent to the drug, was unable to supply it in high enough quantities.

¹¹⁰ Copyright Designs and Patents Act 1988, s30A.

¹¹¹ *ibid* s30.

¹¹² *Pro Sieben Media AG v Carlton UK Television Ltd* [1999] 1 WLR 605, [1999] FSR 610.

¹¹³ Ellen't Hoen, ‘Covid-19 and the Comeback of Compulsory Licensing’ (*Medicines Law & Policy*, 23 March 2020)

<<https://medicineslawandpolicy.org/2020/03/covid-19-and-the-come-back-of-compulsory-licensing/>> accessed 10 February 2023 .

A second consequence of viewing the public domain as being in opposition to the intellectual property regime as the territorial metaphor prompts us to do is that it legitimises the current distribution of intellectual property rights. In other words, it assumes that everything that is in the public domain should remain there in order to leave the resource open to all. However, this is problematic insofar as it masks how this construction of the public domain ‘disadvantages subordinate indigenous and other disempowered groups globally’.¹¹⁴ In particular, this binary rhetoric of ‘intellectual property versus the public domain’ disguises how the current public domain functions more ‘in the interests of traditional property holders than in the interests of the commoners’.¹¹⁵ It does so by suggesting that the public domain is open to free and equal exploitation by all, when in reality those that possess more knowledge and wealth have a higher chance of being awarded with intellectual property rights.

The neem tree incident provides one such illustration of how traditional knowledge that is in the public domain has functioned in favour of the Western world. Indigenous to India, the neem tree has been traditionally used for many medicinal and agricultural purposes. In the 1980s, various Western companies sought to obtain patents for applications of neem extracts. Though ‘Indian enterprises were aware of the commercial value of the tree’, they were ‘unable to invest the resources to patent its derivatives throughout the world’.¹¹⁶ This is an example of how traditional knowledge that rests in the public domain is exploited by those who have the means to convert the knowledge into patentable products. The Stevia plant offers a second example of this: a recent publication named *The Bitter Sweet Taste of Stevia* demonstrated how the Guarani tribe of Brazil who discovered Stevia’s sweetening properties are not receiving a fair or equitable share of the benefits that have arisen from the commercialisation of Steviol glycosides.¹¹⁷

¹¹⁴ Chander and Sunder (n 1) 1335.

¹¹⁵ *ibid* 1343.

¹¹⁶ *ibid* 1353.

¹¹⁷ Public Eye, ‘The Bitter Sweet Taste of Stevia’ (*Public Eye*, 2015)

<<https://www.publiceye.ch/en/publications/detail/the-bitter-sweet-taste-of-stevia>>
accessed 10 February 2023.

Furthermore, the patent system's formal way of recognising knowledge also poses a barrier to members of indigenous communities trying to gain property rights for their traditional knowledge. The patent system only grants patents to inventions that are new, involve an inventive step, and are capable of industrial application. Whether an invention is new is determined with reference to whether an invention is part of the state of the art, determined with reference to prior art. However, traditional knowledge is not 'new', which prevents it from passing the threshold for inventions that most patent laws have. Similarly, the prior art requirement poses a high hurdle for those looking to gain intellectual property rights to their traditional knowledge.

Therefore, in this context too, the binary rhetoric of the public domain ensued from the territorial metaphor poses a barrier to the understanding that the intellectual property regime favours certain kinds of knowledge over others. This inherently disadvantages indigenous and disempowered communities and has resulted in calls to rethink the way the system operates.¹¹⁸

V. THE TERRITORIAL METAPHOR AND WHAT INTELLECTUAL PROPERTY LAW OUGHT TO BE

Thus far, the consequences of the four characteristics associated with the territorial metaphor have been described and discussed, and it has been shown how each misrepresents the public domain and forms a barrier to determining what the intellectual property regime is. The territorial metaphor perpetuates a vision of the public domain that is static, clearly bordered, in binary opposition to intellectual property, and which propels the conviction that propertization incentivises creativity. However, this vision masks several important facts. First, that protecting the public domain through propertization will lead to its underuse, rather than depletion, which will stifle creation rather than encourage it. Second, that there is no clear and tangible border around the public domain which can be invoked to protect it. Third, because the public domain is dynamic and relational, we have the responsibility of safeguarding the balance between the public domain and the intellectual property regime so that both can equally thrive. Fourth, that we cannot overcome the dualism between the public domain and the intellectual property regime as long as we view their

¹¹⁸ Ruth L Okediji, 'Traditional Knowledge and the Public Domain' (2018) 176 CIGI Papers 1, 14.

relationship as binary or mutually exclusive.

Therefore, it is submitted that the metaphorically-driven conception of the public domain as territory fails to recognise the inherently complex and interdependent relationship between the public domain and intellectual property law. So, why is it still being referred to with such weight throughout academic and judicial discussions?

Cohen answers this by saying that the territorial metaphor is still being employed because it helps to justify and normalise the increasing enclosure of the public domain.¹¹⁹ For instance, though anti-commodificationists may argue that the narrowing scope of copyright exemptions will ‘disrupt the balance between the proprietary and the public’¹²⁰ (since they believe these exemptions form part of the public domain), pro-commodificationists can easily dismiss this argument by stating that the public domain is territorially separate from the proprietary. They can use this to argue that narrower copyright exemptions will not affect the public domain because copyrighted works themselves are not *in* the public domain. This is an example of how our conception of the public domain determines our perception of what the scope of the intellectual property system should be.

Unfortunately, the very fact that ‘opponents of the increasing encroachment of the public domain, are using the same metaphor that helps the proponents of a rising commodification in intellectual property to deny that threat, makes it impossible to reach an effective solution for the preservation of the public domain’.¹²¹ Indeed, the term ‘public domain’ alone is so entrenched in denotative and connotative meanings ‘that constitute the artistic, intellectual, and informational public domain as a geographically separate place’,¹²² that it is hard to depart from the metaphor in order to illustrate that the public domain ought to be ‘a varying bundle of rights revolving around the right to access and use’.¹²³

¹¹⁹ Cohen (n 3) 123.

¹²⁰ *ibid.*

¹²¹ Dusollier (n 36) 5.

¹²² Cohen (n 3) 123.

¹²³ Chander and Sunder (n 1) 1339.

Thus, if we adopt the view that intellectual property law ought to better protect the public domain as the building blocks of our culture and creation, then we need to shift away from the territorial metaphor in order to propel this view. Moreover, we should step away from the use of metaphors entirely, because each metaphor has its own set of connotations and characteristics which limit our thinking and misrepresent the public domain. The metaphor of the public domain as a natural resource, for instance, implies that the public domain is depletable. The metaphor of the public domain as a wasteland implies that it should not be the subject of protection. The metaphor of the public domain as an environmental commons ends up reinforcing the 'crucial binary opposition that intellectual property posits between nature and society, and discovery and artifact'.¹²⁴

Therefore, in answering the question of what intellectual property law ought to be, five interdependent suggestions are revealed over the course of this article:

1. Intellectual property law ought to recognise that it is intrinsically and unequivocally related to the public domain.
2. Intellectual property law ought to safeguard the rights and well-being of the public domain as effectively as it safeguards the rights of creators, inventors and innovators.
3. Intellectual property law ought to proactively engage in growing the space of the public domain commons, for the public domain represents the common heritage of humanity.
4. Intellectual property law ought to recognise that the mechanism of growing a commons of knowledge and ideas depends on securing the abundance of free inflow of facts and ideas.
5. Intellectual property law ought to recognise and proactively seek ways to counterbalance the inequalities it creates not only among existing intellectual creators, but also among future creators.

It is argued that we need to first positively and proactively define the scope and substance of the public domain free from any metaphors, in order to then give effect to what the intellectual property regime ought to be. By elevating the discussion of the public domain beyond the use of metaphors, we

¹²⁴ Mario Biagioli, 'Nature and the Commons: The Vegetable Roots of Intellectual Property' (2009) 382 *Max Planck Institute for the History of Science Preprint* 241, 242.

can begin to draw its blueprint in a way that does justice to its importance.

CONCLUSION

This article began with the statement that intellectual property and the public domain are ‘paired together in a perpetual dance’.¹²⁵ This dance has been examined from the viewpoint of thinking about the public domain as a land. It has been shown that using the territorial metaphor to understand what the public domain is clearly acts as a barrier to determining what intellectual property law is.

The metaphor of land perpetuates a vision of the public domain that is static, clearly bordered, and in binary opposition to intellectual property. Furthermore, it propels the conviction that propertization incentivises creativity. Although this vision is easy to understand and experience, it is false insofar as it misrepresents not only what the public domain is, but also what intellectual property law is. Namely, that intellectual property law rests on disputed justifications, with boundaries that are easy to expand and constantly shifting. This negatively impacts its dynamic, complex, and interdependent relationship with the public domain.

In answering the question of what intellectual property law ought to be, it was shown how the vision created by the territorial metaphor has been used by pro-commodificationists to further their agenda of expanding the intellectual property regime at the expense of the public domain. This illustrates how our conception of the public domain determines what we believe the scope of the intellectual property regime should be. Because this territorial conception of the public domain has become so deeply entrenched in the intellectual property debate, those advocating against the expansion of the intellectual property regime have been limited in their attempt to do so because they too have been constrained by the experiences generated by the characteristics of the territorial metaphor.

It was then argued that if we want to reverse this paradigm, we should look to the five suggestions of what intellectual property law ought to be that have been revealed by the discussion in this paper. We need to rise above the use

¹²⁵ Chander and Sunder (n 1) 1339.

of metaphors and positively define the scope and substance of the public domain, rather than perpetuate the current debate. We should look at the role the public domain needs to play in a democratic society and determine intellectual property law accordingly. Intellectual property law must not only assign ownership rights to those who are deserving, but also simultaneously promote cultural heritage, diversity, and the freedom to create. We need to overcome the dualism in our thinking of intellectual property and the public domain, in order to realise that because one depends on the other, we need to equally protect both.

Best of both worlds? The interplay between international human rights law and the law of armed conflict in cyberspace

Tatjana Grote*

ABSTRACT

Cyberoperations are an increasingly prevalent part of armed conflict, as evidenced by the recent cyber-based incidents in Ukraine. While much has been written about how the law of armed conflict (LOAC) applies to such operations, relatively little attention has been paid to the role of international human rights law (IHRL). This article attempts to fill this gap, focusing on the law of international armed conflict. In its first part, it will be demonstrated that essential parts of LOAC are premised on the idea that only operations with violent effects like death, injury or destruction, are of a severity that would justify an obligation to distinguish between civilians and military objectives. As political, economic, and social activities move into the non-physical sphere of cyberspace, this assumption is no longer tenable. Yet, it is averred that simply applying the principle of distinction to all cyberoperations does not constitute a satisfying solution. The article hence turns to IHRL and finds that its focus on individual rights and accountability could compensate for LOAC's shortcomings. However, IHRL's protective value is greatly circumscribed by its limited extraterritorial applicability. The novel conclusion offered in the final section is that the true source of the existing gap in legal protection against military cyberoperations is the failure to acknowledge the possibility of affecting the enjoyment of immaterial goods and freedoms without having physical control over a person. This makes the norms of LOAC inadequate and IHRL inapplicable, leaving civilians in situations of armed conflict vulnerable to severely harmful cyberoperations.

*LLM (LSE) '22, Carlo-Schmid Fellow at the International Criminal Court (ICC). The views expressed by the author in this article are their own and do not represent the views of the ICC.

INTRODUCTION

Recent cyberoperations against government websites and private companies in Ukraine have once again shown that cyberspace plays an important role in modern armed conflict.¹ Long before Russia's invasion of Ukraine, military cyberoperations have been a cause of concern, *inter alia*, for the legal protection of civilians from the effects of armed conflict. Although there now seems to be widespread consensus that the law of armed conflict (LOAC) generally applies to cyberoperations with a nexus to armed conflict,² many commentators lament a lack of protection against one specific type of operation: cyberoperations without any reasonably foreseeable *physical* effects.³ This category could include operations deleting civilian banking data, disabling navigation systems or blocking access to messaging services.

¹ Before the Russian invasion, a malware deleting data was detected in more than a thousand local computer networks. Furthermore, several Distributed Denial of Service Attacks (DDoS) rendered government and banking websites temporarily dysfunctional. On 24 February 2022, cyberoperations against the US firm *Viasat* caused the shutdown of several thousand internet modems, which resulted in a 'huge loss in communications'; James Pearson and Christopher Bing, 'The Cyber War between Ukraine and Russia: An Overview' (*Reuters*, 10 May 2022) <<https://www.reuters.com/world/europe/factbox-the-cyber-war-between-ukraine-russia-2022-05-10/>> accessed 16 February 2023. Ukraine reacted by forming an 'IT army' and launching DDoS operations against the official website of the Kremlin and the office of President Putin. James Pearson, 'Ukraine Launches "IT Army," Takes Aim at Russian Cyberspace' (*Reuters*, 27 February 2022) <<https://www.reuters.com/world/europe/ukraine-launches-it-army-takes-aim-russian-cyberspace-2022-02-26/>> accessed 16 February 2023. All online sources last accessed on 31 January 2023.

² Karine Bannelier, 'Is the Principle of Distinction Still Relevant in Cyberwarfare? From Doctrinal Discourse to States Practice' in Nicholas Tsagourias and Russell Buchan (eds), *Research Handbook on International Law and Cyberspace* (Edward Elgar Publishing 2021) 428; NATO Cooperative Cyber Defence Centre of Excellence (ed), *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations* (Second edition, Cambridge University Press 2017) 375.

³ Cyberoperations can have complex indirect effects. While there is no absolute clarity as to if and to what extent these effects need to be taken into account, this work will adopt the majority approach of the Tallinn Manual and consider a cyberoperations as devoid of physical effects if no physical effects were intended or reasonably foreseeable. NATO Cooperative Cyber Defence Centre of Excellence (n 2) 416, para. 5.

Traditionally, a crucial avenue for providing protection has been the principle of distinction, which obliges belligerent parties to distinguish between civilians and civilian objects on the one hand, and combatants as well as military objectives on the other.⁴ Only the latter two categories constitute legitimate targets. The most important provisions operationalising the cardinal principle of distinction are predominantly framed with reference to ‘attacks’ and ‘objects’. Since these terms have traditionally been interpreted to require physicality in the effects and targets of an operation,⁵ many hold that if an operation does not produce physical effects or damage to *tangible* objects,⁶ it does not qualify as an ‘attack’ (defined in Article 49(1) Additional Protocol 1, AP1). Accordingly, there would be no general requirement to limit such operations to military objectives. Many conclude that this causes a lack of legal protection from severely harmful, non-physical effects of cyberoperations.⁷ While much has been written on whether and to what extent this focus on physical effects and objects limits the applicability of the principle of distinction,⁸ this paper will take a step back and challenge a recurring, underlying assumption of these debates: that the current gap in legal protection of civilians from harmful cyberoperations without physical effects is caused by the aforementioned focus on physicality (‘physical anchoring’)⁹ of LOAC.

⁴ See Article 48 Additional Protocol (AP) 1.

⁵ Traditionally, ‘violence’ has been understood an action producing physical effects; Michael Bothe, Karl Josef Partsch and Waldemar A Solf, *New Rules for Victims of Armed Conflicts: Commentary on the Two 1977 Protocols Additional to the Geneva Conventions of 1949* (Martinus Nijhoff Publishers 1982) 289.

⁶ As will be discussed below, the ICRC commentary as well as many scholars consider ‘tangibility’ an inherent characteristic of any ‘object’.

⁷ The phrase ‘cyber-contained’ will be used to denote cyberoperations without any reasonably foreseeable physical effects.

⁸ This perspective is especially relevant with a view to the hybrid model of warfare, which combines conventional methods of warfare with disinformation operations and economic measures. In times where what is traditionally understood as violence – physical force – is combined with other types of force, clarifying which norms of international law are applicable to which subsets of operations forming part of this form of warfare is crucial. On hybrid warfare, see Arsalan Bilal, ‘NATO Review – Hybrid Warfare – New Threats, Complexity, and “Trust” as the Antidote’ (*NATO Review*, 30 November 2021)

<<https://www.nato.int/docu/review/articles/2021/11/30/hybrid-warfare-new-threats-complexity-and-trust-as-the-antidote/index.html>> accessed 16 February 2023.

⁹ Henning Lahmann, ‘Protecting the Global Information Space in Times of Armed Conflict’ (2020) 102 *International Review of the Red Cross* 1227, 1245.

In doing so, this paper will only examine cyberoperations with a nexus to armed conflict, i.e., operations to which LOAC is generally applicable.¹⁰ Such operations are considered most likely to be conducted by states in support of kinetic operations in an international armed conflict (IAC).¹¹ Furthermore, it should be noted that operations with a substantive impact on the civilian population are often traced back to state-led cyber commands as they require a high level of technical expertise and resources, which many non-state actors lack.¹² Therefore, the following analysis will focus on the law of IAC from a state actor perspective. While the issue of attribution (i.e., legally connecting the actions of private persons to a state) is highly relevant, it is less crucial for this work since it is not focused on ensuring accountability, but on the rules limiting states' conduct in cyberspace in the first place.¹³ Moreover, attribution is usually less problematic if an operation is linked to ongoing, kinetic hostilities.¹⁴ This was illustrated by the attribution of a cyberoperation by, *inter alia*, the European Union, to Russia at the dawn of its invasion of Ukraine.¹⁵

¹⁰ See Tallinn Manual (n 2), Rule 80; NATO Cooperative Cyber Defence Centre of Excellence (n 2) 376, para. 5.

¹¹ Kinetic operations are understood as operations relying on the motion of physical material bodies and the associated release of energy to create physical effects. An example of such an operation would be the destruction of a bridge using a bomb. Terry D Gill, 'International Humanitarian Law Applied to Cyber-Warfare: Precautions, Proportionality and the Notion of "Attack" under the Humanitarian Law of Armed Conflict' in Nicholas Tsagourias and Russell Buchan (eds), *Research Handbook on International Law and Cyberspace* (Edward Elgar Publishing 2021) 458; *ibid* 469; 'Definition of Kinetic' (*Merriam-Webster*, 30 December 2022) <<https://www.merriam-webster.com/dictionary/kinetic>> accessed 16 February 2023.

¹² Richard Stiennon, 'A Short History of Cyber Warfare' in James A Green (ed), *Cyber warfare: a multidisciplinary analysis* (Routledge 2015) 14.

¹³ On the differing relevance of the issue of attribution, see Laurent Gisel, Tilman Rodenhäuser and Knut Dörmann, 'Twenty Years on: International Humanitarian Law and the Protection of Civilians against the Effects of Cyber Operations during Armed Conflicts' (2020) *International Review of the Red Cross* 287, 309.

¹⁴ James Gow, 'The Ambiguities of Cyber Security' in James Gow and others (eds), *Routledge Handbook of War, Law and Technology* (1st edn, Routledge 2019) 120.

¹⁵ Council of the EU, 'Russian Cyber Operations against Ukraine: Declaration by the High Representative on Behalf of the European Union' (*Council of EU*, 10 May 2022) <<https://www.consilium.europa.eu/en/press/press-releases/2022/05/10/russian-cyber-operations-against-ukraine-declaration-by-the-high-representative-on-behalf-of-the-european-union/>> accessed 16 February 2023.

The following section will give an overview of military cyberoperations and the existing LOAC literature thereon. Section Three will then argue that, even if applicable, it seems unlikely that the principle of distinction can provide meaningful protection against non-physical harm as the interconnected and complex nature of cyberspace make both the distinction and assessment of proportionality an overwhelmingly difficult exercise. Section Four will then illustrate why international human rights law (IHRL) might be more apt to regulate cyberoperations meaningfully due to its focus on transparency, accountability and individual rights. However, a crucial shortcoming of the protection provided by IHRL remains its limited extraterritorial applicability. Taking both sections together, it is concluded that the current gap in legal protection arises from more than just the non-violent nature of certain cyberoperations. Both LOAC and IHRL provide protection against non-physical harm, but only for persons in the hands of, or under the effective control of, a state. Hence, the current lack of protection is not solely attributable to LOAC's focus on physical effects and objects, but rather to the possibility of *directly* targeting immaterial goods and values *without having physical control* over the persons who will feel the consequences of such operations.

I. 'UPDATE REQUIRED'? WHY SOME MILITARY CYBEROPERATIONS MIGHT NOT BE COVERED BY THE PRINCIPLE OF DISTINCTION

Cyberoperations have been defined by the International Committee of the Red Cross (ICRC) as 'operations against a computer, a computer system or network, or another connected device, through a data stream'.¹⁶ Operations forming part of this category can produce idiosyncratic effects on a great variety of targets. Due to this peculiarity, it is crucial to differentiate between different types of operations as some are accepted to be covered by existing law, whereas others are considered to fall between the gaps of the current LOAC. This section will describe some of the most notable cyber-incidents and elaborate on

¹⁶ International Committee of the Red Cross, 'International Humanitarian Law and Cyber Operations during Armed Conflicts: ICRC Position Paper Submitted to the Open-Ended Working Group on Developments in the Field of Information and Telecommunications in the Context of International Security and the Group of Governmental Experts on Advancing Responsible State Behaviour in Cyberspace in the Context of International Security, November 2019' (2020) 102 *International Review of the Red Cross* 481, footnote 1.

why some of these operations might elude the current protections under LOAC. In discussing how commentators propose to address the identified gap in legal protection, it will be argued that the existing proposals remain vague and overall unconvincing.

A preliminary note on the scope of the following analysis is warranted. As mentioned above, this article focuses on the law of international armed conflict from a state actor perspective. This focus has been chosen for the following reasons. Many of the most consequential operations have been tentatively linked to state actors.¹⁷ This must not be confused with a legal attribution of such operations to a state. Yet, it shows that despite the certainly crucial role of non-state actors in cyberspace, state actors are still at least as relevant. In fact, as states have ‘inexhaustible resources, an educated and trained workforce of cyber-attackers, and military and economic motivations’,¹⁸ it might generally be more likely that large-scale operations with a significant impact on the civilian population are conducted by states rather than non-state actors.¹⁹ Especially in armed conflict, where quick and precise action is essential, it seems unlikely that the many states which have established military cyber commands over the past years²⁰ would completely outsource the most impactful (and hence sophisticated) operations to non-state groups. While questions of attribution and the obligations of non-state actors remain highly relevant, they are beyond the remit of this work since the focus here is on the obligations to which states have to adhere when conducting cyberoperations. This is not to say that holding states accountable for their cyberoperations, which requires attribution, is not pivotal. Rather, this article deliberately focuses on the question preceding attribution for the purpose of holding states accountable with respect to their obligations under international law, namely: what are these obligations?

¹⁷ Most notably *Stuxnet* and the recent Ukraine satellite internet incident, which will be explored below.

¹⁸ Stiennon (n 12) 14.

¹⁹ Gill (n 11) 427.

²⁰ Stiennon (n 12) 25–27.

1.1 Existing protections: which cyberoperations are covered by the law of targeting as it stands?

When cyberoperations have reasonably foreseeable physical effects, the law of targeting (i.e., those norms regulating which persons and objects can lawfully be targeted in armed conflict) applies as usual. There is widespread agreement that an operation with a nexus to armed conflict intended to physically damage, kill or cause injury can be considered an ‘attack’. Note that the reference to ‘violence’ in Article 49(1) AP1 does not relate to the means of creating certain effects, but the effects as such.²¹ Hence, if violent effects (including serious mental injury)²² are created through non-violent cyberoperations, they nevertheless qualify as attacks. One example of such an operation is the malware *Stuxnet*, which significantly damaged turbines in the Iranian nuclear facility Natanz, thus delaying Iran’s nuclear program by roughly two years.²³ The operation was significant because it was one of the few operations to this date that employed cyber-means to produce physical damage,²⁴ and which could hence be qualified as an attack within the meaning of Article 49(1) AP1.

However, modern reliance on cyberspace introduces a critical new harm vector: by targeting sensitive data or rendering crucial cyber-based services dysfunctional or inaccessible, severe harm can be caused, even in absence of physical effects. This was shown by several ransomware operations affecting health facilities, *inter alia*, in the UK.²⁵ Yet, granted a nexus with armed conflict exists, such operations are covered by extant LOAC since certain targets like

²¹ NATO Cooperative Cyber Defence Centre of Excellence (n 2) 415.

²² Eian Katz, ‘Liar’s War: Protecting Civilians from Disinformation during Armed Conflict’ (2020) 102 International Review of the Red Cross 659, 661.

²³ Stiennon (n 12) 20.

²⁴ Danny Steed, ‘The Strategic Implications of Cyber Warfare’ in James A Green (ed), *Cyber warfare: a multidisciplinary analysis* (Routledge 2015) 81.

²⁵ *WannaCry* resulted in the cancellation of at least 19,497 medical appointments and operations. Staff in hospitals and GP practices were locked out of their devices in 24 acute trust, leading to emergency patients having to be diverted and moved to other hospitals. Comptroller and Auditor General, ‘Investigation WannaCry Cyber Attack and the NHS (Summary)’ (UK National Audit Office 2018) Session 2017–2019 HC 414 7–8 <<https://www.nao.org.uk/wp-content/uploads/2017/10/Investigation-WannaCry-cyber-attack-and-the-NHS-Summary.pdf>> accessed 16 February 2023; Tim McCormack, ‘International Humanitarian Law and the Targeting of Data’ (2018) 94 International Law Studies 222, 236.

health facilities,²⁶ humanitarian relief operations,²⁷ cultural objects and places of worship²⁸ as well as objects indispensable to the survival of the civilian population²⁹ are protected not only against attacks, but against ‘military operations’ or, for cultural objects and places of worship, any ‘act of hostility’.³⁰ Nevertheless, these operations constitute only a small subset of all possible operations with severely harmful effects on civilians.

More general protection is provided by Article 57(1) AP1, which stipulates that the belligerents have to take constant care to spare the civilian population during all operations.³¹ However, this is a rather vague standard.³² It is only the more specific rules of the principle of distinction that allow states to be held accountable with respect to concrete obligations and prohibitions.

1.2 Gaps in protection: how to undo the physical anchoring of the principle of distinction?

While the previous subsection illustrated that certain cyberoperations are satisfyingly covered by existing law, the following section will illustrate why those operations not creating any physical effects against targets other than those enjoying special protection might escape the reach of the specific prohibitions and obligations of the principle of distinction.

²⁶ Article 12(1) AP1.

²⁷ Articles 70(4), 71(2) AP1.

²⁸ Article 53(a) AP1.

²⁹ Article 54(2) AP1.

³⁰ This does not mean that this protection is absolute. In fact, the protection of cultural property and objects indispensable to the survival of the civilian population are caveated by exceptions based on the concept of military necessity (see Article 4(2) Hague Convention for the Protection of Cultural Property, Article 56(5) AP1). Gisel, Rodenhäuser and Dörmann (n 13) 327–329.

³¹ Eric Talbot Jensen, ‘Cyber Attacks: Proportionality and Precautions in Attack’ (2013) 89 *International Law Studies* 198, 202–203.

³² This standard is less strong than the more specific rules in Articles 51–56 AP1 since it does not constitute a prohibition on targeting certain objects or persons. Instead of defining explicitly what belligerent parties must or must not do, Article 57(1) AP1 leaves it to commanders to define which steps are necessary and feasible to take constant care. Concrete duties are only spelled out in the subsequent paragraphs, which refer to ‘attacks’, not military operations. Overall, Article 57(1) has been compared to Article 48 AP1, since both contain general principles which are only clarified by other provisions.

The principle of distinction is formulated in general terms in Article 48 AP1 with reference to all ‘military operations’:

In order to ensure respect for and protection of the civilian population and civilian objects, the Parties to the conflict shall at all times distinguish between the civilian population and combatants and between civilian objects and military objectives and accordingly shall direct their operations only against military objectives.

However, provisions operationalising the principle only mention ‘attacks’.³³ Therefore, the interpretation of its definition (‘acts of violence’, Article 49(1) AP1) has been considered crucially important. Some commentators propose a *strict physical effects approach* according to which only a cyberoperation intentionally or reasonably foreseeably causing violent effects in a traditional sense, i.e., physical effects, would qualify as an attack.³⁴ While this approach stays true to the ordinary meaning of the term ‘violence’ at the time when the drafters chose this term,³⁵ it leads to absurd results: the physical destruction of a server hosting a critical civilian service would be unlawful while rendering it dysfunctional through cyber-means (e.g., like the cyberoperations against Ukraine in 2022) would be lawful despite producing the same effect on civilians.³⁶ Apart from this conceptually unsatisfying outcome, following the strict physical effects approach would substantially limit the legal protection against cyberoperations provided by the principle of distinction, potentially leading to a situation where an operation severely harming civilians would be perfectly lawful under LOAC.³⁷

³³ See, e.g., Article 51(2), (4), (5), (6); Article 52(2), Article 55(2), Article 56(1), Article 57(2) AP1.

³⁴ Yoram Dinstein, ‘The Principle of Distinction and Cyber War in International Armed Conflicts’ (2012) 17 *Journal of Conflict and Security Law* 261, 264.

³⁵ Kubo Mačák, ‘Military Objectives 2.0: The Case for Interpreting Computer Data as Objects under International Humanitarian Law’ (2015) 48 *Israel Law Review* 55, 67.

³⁶ Noam Lubell, ‘Lawful Targets in Cyber Operations: Does the Principle of Distinction Apply?’ (2013) 89 *US Naval War College International Law Studies* 252, 266.

³⁷ Cordula Droege, ‘Get off My Cloud: Cyber Warfare, International Humanitarian Law, and the Protection of Civilians’ (2012) 94 *International Review of the Red Cross* 533, 561.

The consequences of adopting a *strict physical effects approach* are of great relevance in light of past cyberoperations. In 2007, Estonia became the target of one of the first large-scale cyberoperations. In reaction to the removal of a Soviet statue, Estonian banking, telecommunications and government websites were flooded with requests from hijacked computers,³⁸ causing many services to be unavailable for some time. One year later, Georgia became the target of similar operations which were considered to be linked with the Russian use of force in the same year.³⁹ First, hijacked computers overwhelmed government-related cyber-services, including the Social Assistance and Employment State agency, with fake requests thus causing them to break down.⁴⁰ Then, on the eve of the Russian invasion of Georgia, essential cyber-based communication systems were disabled, effectively isolating the country and its citizens from their government and the rest of the world.⁴¹

More recently, military cyberoperations were employed in connection with the armed conflict in Ukraine. On the eve of the Russian invasion in 2022, a cyberoperation rendered thousands of satellite internet modems dysfunctional, ‘causing indiscriminate communication outages and disruptions across several public authorities, businesses and [private] users in Ukraine’.⁴² Furthermore, multiple distributed denial-of-service (DDoS) operations crippled the websites of several Ukrainian banks and government agencies.⁴³

All three examples highlight that cyberoperations can cause highly disruptive, and yet entirely non-physical, effects. If the principle of distinction does not apply in such operations, civilians might lawfully be subjected to such

³⁸ Stiennon (n 12) 17. This type of operation is called Distributed-Denial-of-Service (DDoS) operation and is one of the less sophisticated but common ways to render services dysfunctional; UK National Cybersecurity Centre, ‘Denial of Service (DoS) Guidance’ (*National Cyber Security Centre*, 16 March 2016) <<https://www.ncsc.gov.uk/collection/denial-service-dos-guidance-collection>> accessed 16 February 2023.

³⁹ Stiennon (n 12) 20.

⁴⁰ *ibid* 18.

⁴¹ *ibid* 19.

⁴² Council of the EU (n 15).

⁴³ UK Government, ‘UK Assesses Russian Involvement in Cyber Attacks on Ukraine’ (*GOV.UK*, 18 February 2022) <<https://www.gov.uk/government/news/uk-assess-russian-involvement-in-cyber-attack-s-on-ukraine>> accessed 16 February 2023.

effects under IHL. In light of this, three avenues have been explored to undo this ‘physical anchoring’ of the principle of distinction.

2.2.1 Option 1: *The principle of distinction applies to all military operations*

Some commentators contend that the principle of distinction applies not only to attacks, but to military operations more broadly. According to one commentator, the principles of distinction, proportionality and precaution,⁴⁴ ‘accurately understood’, apply to *hostilities*, i.e., all operations that are meant to harm the adversary through physical effects or other means.⁴⁵ Based on a systematic interpretation of AP1, another commentator avers that if the principles of distinction, proportionality and precaution only applied to attacks, Article 48 AP1 in general, and the *chapeaux* of Article 51 and Article 57 AP1, would be superfluous.⁴⁶

These views, however, are unconvincing. In fact, it is hard to deny that the more specific norms of LOAC, which are considered to give effect to the principle of distinction,⁴⁷ explicitly refer to attacks. During the negotiations of AP1, the delegation of the UK, in fact, pointed out that the broadly framed basic rule in Article 48 AP1 was meant as a general rule and commanders would rather consult Article 52 AP1, which contains a specific prohibition on targeting civilian objects and a definition thereof, when making targeting decisions.⁴⁸ The

⁴⁴ The principle of proportionality can be found in Article 51(5)(b) AP1 and stipulates that the incidental civilian deaths, injuries and damage to civilian objects caused by an operation must not be excessive in relation to its anticipated military advantage. The principle of precaution set out in Articles 57 and 58 AP1 and requires belligerent parties to take certain measures with a view to minimising incidental civilian deaths, injury and destruction to civilian objects as well as to prevent disproportionate attacks.

⁴⁵ Nils Melzer, ‘Cyberwarfare and International Law’ (2011) 27 UNIDIR <<http://unidir.org/files/publications/pdfs/cyberwarfare-and-international-law-382.pdf>> accessed 16 February 2023.

⁴⁶ Heather H Dinness, *Cyber Warfare and the Laws of War* (Cambridge University Press 2012) 196–202.

⁴⁷ Michael N Schmitt, ‘“Attack” as a Term of Art in International Law: The Cyber Operations Context’ (2012) 4th International Conference on Cyber Conflict 7.

⁴⁸ Diplomatic Conference on the Reaffirmation and Development of International Humanitarian Law Applicable in Armed Conflicts Geneva, *Official Records of the Diplomatic Conference on the Reaffirmation and Development of International Humanitarian Law Applicable in Armed Conflicts, Geneva (1974-1977)*, vol XIV (Federal Political Department 1978) 28, para 19. In light of the persistent unclarity and potentially absurd conclusions when using the

delegation of Israel explicitly stated that the reference to 'military operations' in Article 51(1) AP1 should not be seen as a derogation from the narrower protection of the civilian population against attacks only, and suggested the term ('attack') instead to avoid inconsistencies.⁴⁹ These statements are hard to ignore, especially because they are in line with current state practice.⁵⁰

In light of the foregoing, this author believes that the use of the term 'military operations' in the *chapeaux* of Articles 51 and 57 as well as Article 48 AP1 does not allow for the conclusion that specific rules which solely mention the term 'attack' are applicable to all 'military operations'. Indeed, it would be hard to explain why the drafters felt the need to legally define 'attacks' instead of 'military operations', if all the provisions using this term were applicable to the latter category instead of the former. Instead, Articles 51 and 57 as well as Article 48 AP1 should be considered introductory, preamble-like provisions, which formulate the general principle underlying the subsequent rules without altering the content of these more specific norms.

Moreover, the precise definitions of 'hostilities' and 'military operations' remain unclear. The first is notoriously controversial⁵¹ and the latter is defined in two different ways by the ICRC commentary to AP1.⁵² Hence, even if the principles of distinction, proportionality and precaution did apply to all military operations or hostilities, there is still a need to delimit these two categories from another category of military activity which can lawfully be directed at civilians. Such a category undoubtedly exists given that espionage and

means of interpretation mentioned in Article 31 Vienna Convention of the Law of Treaties, a recourse to the preparatory work under Article 32 VCLT seems warranted.

⁴⁹ *ibid* 223, paras 36-37.

⁵⁰ Schmitt (n 47) 7.

⁵¹ The notion of 'hostilities' has been relevant in the context of defining the phrase 'direct participation in hostilities'. This phrase is of utmost important as civilians directly participating in hostilities can be targeted lawfully (see Article 51(3) AP1, Article 13(3) Additional Protocol 2); Nils Melzer and International Committee of the Red Cross, *Interpretive Guidance on the Notion of Direct Participation in Hostilities under International Humanitarian Law* (International Committee of the Red Cross 2009) 43.

⁵² Compare Claude Pilloud and others (eds), *Commentary on the Additional Protocols of 8 June 1977 to the Geneva Conventions of 12 August 1949* (International Committee of the Red Cross: Martinus Nijhoff Publishers; Distributors for the United States and Canada, Kluwer Academic Publishers 1987) 617, para 1936 with *ibid* 680, para 2191.

psychological operations are widely accepted as lawful, even where conducted indiscriminately.⁵³

2.2.2 Option 2: Re-interpreting the term 'violence'

Another avenue for making all rules of targeting applicable to cyberoperations without physical effects is to reinterpret the meaning of 'violence' in Article 49(1) AP1. The experts of the Tallinn Manual (TM)⁵⁴ introduced the *functionality approach* by defining a cyberattack as 'a cyber operation ... that is reasonably expected to cause injury or death to persons or damage or destruction to objects'.⁵⁵ At first sight, the definition does not seem to deviate from a *strict physical effects approach*. The key innovation is hidden in the TM majority's conclusion that a mere loss of functionality could qualify as 'damage'. However, most experts held that this was only the case if the restoration of functionality would require the replacement of a physical component. This reading hence maintains the requirement of physical effects but covers even *de minimis* damage, where it renders a system dysfunctional.⁵⁶ This insistence on some kind of physical effect is in line with the legal positions expressed in statements of several states, for example Peru and Denmark, on the application of LOAC to cyberoperations.⁵⁷

Only few TM experts contend that a mere loss of functionality, regardless of which type of action is required to render the targeted system or device functional again, would amount to 'damage' within the meaning of AP1.⁵⁸ This view, however, has been prominently expressed outside the TM process. One commentator points to the mention of 'neutralisation' in Article 52(2) AP1 to suggest that the drafters did not consider the notion of attacks to be strictly limited to violent operations, but meant to also include operations

⁵³ Michael N Schmitt, 'Wired Warfare 3.0: Protecting the Civilian Population during Cyber Operations' (2019) 101 *International Review of the Red Cross* 333, 340.

⁵⁴ The Tallinn Manual is an academic publication on the international law applicable to cyber warfare. It is the outcome of discussions between roughly twenty experts of international law. Despite not being binding, it is one of the most commonly referenced and comprehensive publications on international law and military cyberoperations.

⁵⁵ NATO Cooperative Cyber Defence Centre of Excellence (n 2) 415.

⁵⁶ The TM experts held that, otherwise, *de minimis* damage would not be included. *ibid* 416.

⁵⁷ Gisel, Rodenhäuser and Dörmann (n 13) 314.

⁵⁸ NATO Cooperative Cyber Defence Centre of Excellence (n 2) 418.

affecting the functionality of a military objective.⁵⁹ The ICRC recently confirmed its view that ‘during an armed conflict an operation designed to disable a computer or a computer network constitutes an attack under [LOAC]’, regardless of how the system is disabled.⁶⁰ Such a strict functionality-based approach has been adopted by France.⁶¹

Another commentator suggests that if the effects of a cyberoperation exceed merely transient inconveniences and ‘cause significant functional harm to infrastructures’, they should be considered ‘attacks’.⁶² Such a focus on critical infrastructure was also reflected in the positions of Bolivia, Ecuador and Guyana.⁶³ Germany takes a very broad effects-based approach, stating that any ‘harmful effects’ on cyber systems or data amount to an attack.⁶⁴ The criterion of permanency was emphasised by another author.⁶⁵ One commentator suggested that a cyberoperation amounts to an attack if the effects it produces would be considered ‘damage’ if they were the result of a kinetic operation.⁶⁶ Both Australia and the UK subscribe to this idea.⁶⁷

It is therefore clear that there are varied views on whether and how to reinterpret violence. Some commentators focus on the functionality of a system,

⁵⁹ Knut Dörmann, ‘Applicability of the Additional Protocols to Computer Network Attacks’ (International Committee of the Red Cross Resources, 19 November 2004) 4 <<https://www.icrc.org/en/doc/resources/documents/misc/68lg92.htm>> accessed 16 February 2023; Droegge (n 37) 558.

⁶⁰ International Committee of the Red Cross (n 16) 488–489.

⁶¹ Gisel, Rodenhäuser and Dörmann (n 13) 315–316.

⁶² Marco Roscini, ‘Proportionality in Cyber Targeting’ in James Gow and others (eds), *Routledge Handbook of War, Law and Technology* (Routledge 2019) 90.

⁶³ Gisel, Rodenhäuser and Dörmann (n 13) 316.

⁶⁴ Federal Government of Germany, ‘On the Application of International Law in Cyberspace (Position Paper)’ (March 2021) <<https://www.auswaertiges-amt.de/blob/2446304/32e7b2498e10b74fb17204c54665bdf0/on-the-application-of-international-law-in-cyberspace-data.pdf>> accessed 16 February 2023.

⁶⁵ Gill (n 11) 467.

⁶⁶ Jensen (n 31) 205–206.

⁶⁷ Gisel, Rodenhäuser and Dörmann (n 13) 314; United Kingdom Mission to the United Nations, ‘United Nations Group of Governmental Experts on Advancing Responsible State Behaviour in Cyberspace in the Context of International Security - Application of International Law to States’ Conduct in Cyberspace (UK Statement)’ (2021) <https://ccdcoe.org/uploads/2018/10/UK_application-of-international-law-to-states-conduct-in-cyberspace-uk-statement.pdf> accessed 16 February 2023.

others on the severity and permanency of the effects of an operation for civilians, others on an equivalency with the effects of a kinetic attack.⁶⁸ All these approaches share one problematic feature. If physical effects are discarded as a threshold, there is a need to create a straightforward alternative test to identify situations in which the harm to civilians emanating from an operation is so great that a distinction is necessary. Only few would dispute that such a threshold exists as espionage and psychological operations are traditionally not considered to be covered by the principle of distinction.⁶⁹ Even the ICRC stated that civilians are not protected from certain inconveniences caused by armed conflict.⁷⁰

The strict functionality approach sets out a clear threshold, namely at the level where the functionality of a system is impacted. This assumes that there is a strong and consistent correlation between severely harmful effects on civilians and a loss of functionality. However, this assumption might not hold for all systems. For instance, if a belligerent party were to disable a streaming platform, this would arguably not amount to an operation from which civilians need to be shielded. It thus seems reasonable to refer to the severity of the effects of an operation on the civilian population. The above discussion showed that some authors have proposed effects which exceed ‘merely transient inconveniences’ as a new threshold. Yet, the term ‘inconveniences’ seems unhelpful as it is not mentioned in LOAC treaties, and is hence not defined.⁷¹ Equally, the notion of ‘effects equivalent to those of a kinetic attack’ would require further clarification to provide meaningful guidance. Another approach that has been proposed is to accord protection to ‘essential civilian functions or

⁶⁸ Note that one crucial question of all approaches focusing on the effects of an operation is which effects are considered as directly attributable to a specific operation. For instance, an operation leaking sensitive data of certain vulnerable individuals might lead to these persons being harmed. The Tallinn Manual experts held that only if such indirect effects are reasonably foreseeable or intended do they need to be taken into account. This standard seems widely accepted and has not been the subject of much academic discussion. However, the question of how to legally classify indirect effects of cyberoperations should be kept in mind as one that could potentially bring many operations within the scope of the traditional definition of ‘attacks’; NATO Cooperative Cyber Defence Centre of Excellence (n 2) 416, para. 5.

⁶⁹ Michael N Schmitt, ‘Rewired Warfare: Rethinking the Law of Cyber Attack’ (2014) 96 *International Review of the Red Cross* 189, 193.

⁷⁰ Schmitt, ‘Wired Warfare 3.0’ (n 53) 339.

⁷¹ Bannelier (n 2) 441.

services', instead of all types of services and data. However, this would require consensus as to which parts of the ICT infrastructure, and which online services and types of data are essential. While there might be a core set of services and infrastructure commonly accepted as essential, disagreement about the precise extent of this category seems highly likely.⁷²

Consequently, while there are many general ideas on how to re-interpret violence, the devil seems to lie in the details of proposing a stringent and clear alternative threshold for making the principle of distinction applicable. As will be proposed below, there might be merit in the idea of discarding an exclusive focus on LOAC in this respect, and instead consider the human rights guarantees and the relevant jurisprudence thereon when assessing which type and degree of harm civilians must not be subjected to.

2.2.3 Option 3: Data as an object

Another way to bring certain cyberoperations without physical effects under the protective umbrella of the principle of distinction would be to interpret data as an object within the meaning of Article 52(1) AP1, which states that '[c]ivilian objects shall not be the object of attack or of reprisals.' Civilian objects are negatively defined as all objects which do not constitute military objectives as defined in Article 52(2) AP1, i.e., 'objects which by their nature, location, purpose or use make an effective contribution to military action and whose total or partial destruction, capture or neutralisation, in the circumstances ruling at the time, offers a definite military advantage'. While Article 52(1) AP1 also mentions 'attacks', it has been argued that if data was considered an object, its deletion would then clearly constitute damage, and the operation hence an attack.⁷³ While the TM majority rejected this approach with a view to the ordinary meaning of the term 'object' and its definition in the ICRC Commentary to AP1 as 'something that is visible and tangible',⁷⁴ one commentator suggested that this 'ordinary meaning' is not as clear as the TM experts suggested.⁷⁵ Another commentator distinguished between content level data (digital information understandable by humans) and operational level data

⁷² Schmitt, 'Wired Warfare 3.0' (n 53) 346–347.

⁷³ *ibid* 340.

⁷⁴ Pilloud and others (n 52), para 2008.

⁷⁵ Maćák (n 35) 59.

(code determining the functioning of cyber-systems) and claimed the latter could be a military objective.⁷⁶ These approaches are increasingly adopted by states: Finland, Germany, Romania and Norway now hold that civilian data is protected as a civilian object.⁷⁷ France considers civilian content level data protected.⁷⁸

There is some intuitive appeal in both approaches as they would allow for the extension of protection of civilian objects to civilian data. However, two problems emerge.

First, interpreting data as an object alone will not protect civilians from all potentially harmful operations. Imagine a DDoS operation on the banking sector that makes monetary withdrawals impossible. Denial of Service (DoS) operations render systems dysfunctional without destroying data or altering code in any way. Instead, they flood a system with requests, causing it to break down.⁷⁹ Accordingly, this type of operation can create severe consequences without deleting or tampering with any data.⁸⁰ Prohibiting the deletion of civilian data alone thus would not provide protection against the effects of DDoS or similar operations.

Second, granting the same protection to all civilian data might be overinclusive. Arguing that there is a need to extend the general legal protection provided by the principle of distinction to civilian objects to *all* types of data relies on the assumption that the deletion or alteration of any data inevitably causes unacceptable civilian harm (as it is the latter the principle of distinction is

⁷⁶ Heather A Harrison Dinniss, 'The Nature of Objects: Targeting Networks and the Challenge of Defining Cyber Military Objectives' (2015) 48 *Israel Law Review* 39, 41.

⁷⁷ Kubo Mačák, 'Unblurring the Lines: Military Cyber Operations and International Law' (2021) 6 *Journal of Cyber Policy* 411, 421–422. Denmark, Chile and Israel, on the other hand, insist on tangibility.

⁷⁸ *Ministre des Armées Françaises, 'Droit international appliqué aux opérations dans le Cyberspace'* (2019) 14

<<https://www.justsecurity.org/wp-content/uploads/2019/09/droit-internat-appliqu%C3%A9-aux-op%C3%A9rations-cyberspace-france.pdf>> accessed 16 February 2023.

⁷⁹ UK National Cybersecurity Centre (n 38). Note that if a civilian chooses to participate in sending such requests, this could have significant consequences regarding the question of whether they are targetable as civilians directly participating in hostilities. For a detailed discussion of this issue, see David Turns, 'Cyber Warfare and the Notion of Direct Participation in Hostilities' (2012) 17 *Journal of Conflict and Security Law* 279.

⁸⁰ UK National Cybersecurity Centre (n 38).

meant to prevent). Especially when applying the principle of proportionality, a blanket interpretation of data as an object might be problematic since considering every deleted byte of civilian data as damage to a civilian object might quickly make an operation disproportionate even when the actual civilian harm is minimal, which will be explored in more detail below. Therefore, as with the re-interpretation of violence, there is a need to establish guiding criteria on how the deletion or alteration of data translates into civilian harm. Yet, the existing norms of LOAC as well as the academic literature on this issue seem to provide relatively little guidance on this issue.

More generally, all three avenues have one issue in common: they are helpful in the sense that they offer text- and purpose-based reasons for why even operations without violent effects must not be directed at civilians. However, they fall short in providing clear and sensible criteria to identify situations in which the non-violent effects of cyberoperations would translate into unacceptable civilian harm, thus requiring distinction. Before the emergence of cyber-based methods of warfare, ‘violence’ (understood as the creation of physical effects) might have been a reasonably accurate proxy for identifying such situations since (in absence of territorial control) physically destroying material civilian objects or injuring civilians was the only way to cause both direct material and significant immaterial harm to the civilian population. To find a similarly accurate proxy for cyberspace is necessary since cyber-systems and data do not merit humanitarian protection because of their inherent value, but rather due to their relevance to the well-being of the civilian population. However, as illustrated above, this relevance differs substantially between systems and data types,⁸¹ which makes general proxies like a loss of functionality or the deletion of any data seem inadequate. Consequently, a more dynamic framework might be required.

⁸¹ For instance, as discussed above, the value of a streaming service and the associated data clearly differs from that of an essential government service.

II. 'UPDATE FAILED'? ENSURING THE APPLICABILITY OF THE PRINCIPLE OF DISTINCTION IS INSUFFICIENT TO EFFECTIVELY PROTECT CIVILIANS FROM HARM CAUSED BY CYBEROPERATIONS WITHOUT PHYSICAL EFFECTS

Although it was shown that opinions as to which re-adjustments of the existing law would be most desirable and acceptable differ widely, most commentators seem to agree on one underlying assumption: the law of targeting *should* be applicable because it would provide meaningful protection from immaterial harms caused by military cyberoperations during armed conflict. However, the rules of the law of targeting – most importantly the prohibition on attacking civilians⁸² and civilian objects, the prohibition on disproportionate attacks and the obligation to take all feasible precautions – were drafted with conventional land warfare in mind.⁸³ The following section will hence explore if the law of targeting can still provide adequate protection if applied to operations which do not produce reasonably foreseeable physical effects. This will allow an assessment of whether the gap in legal protection alleged by many commentators could, indeed, be remedied by interpreting extant LOAC in a way that would ensure the applicability of the principle of distinction.

2.1.1 Distinguishing between military objectives and civilians objects in an interconnected space

The UK has stated that 'targeting can be challenging in cyberoperations due to the potential dual use nature of some targets, such as infrastructure'.⁸⁴ While the identification of what qualifies as a lawfully targetable military objective under Article 52(2) AP1 has always been difficult, the interconnected nature of cyberspace might exacerbate these difficulties and lead

⁸² As operations without physical effects will, by definition, not injure or kill individuals, the distinction between combatants and civilians will not be considered in the following. Only if an individual's data or online activities were conceived as their virtual self, would a distinction between 'virtual civilians' and 'virtual combatants' become relevant to the debate on cyber-contained operations. However, such an interpretation of cyber-activity and -systems has not taken up a prominent place in the pertinent academic debates.

⁸³ See Article 49(2) AP1.

⁸⁴ Bannelier (n 2) 450.

to a situation 'where nothing civilian remains'.⁸⁵ Imagine a military base sending information to another base. Once the relevant communication packages leave the local network of the sending base, the bits that are being sent will most likely travel through cables and servers which are also used for civilian cyber-systems.⁸⁶ This thereby makes it very difficult to separate civilian and military segments of cyberspace.⁸⁷ To the extent that cyberspace is centralised with one 'upstream' server shifting information back and forth from both military and civilian 'downstream' devices, targeting a central server will offer a military advantage as it will impede communications between the connected military devices. Therefore, the server will qualify as a military objective by use. The acting party does not need to know exactly which routes the information will travel through; it is sufficient that it was reasonably likely that the targeted channel or server was used.⁸⁸ Some commentators hence argue that 'in theory, a country's entire cyber infrastructure could be qualified as a military objective once engaged in armed conflict'.⁸⁹

In some way, this problem can be seen as a corollary of the fact that cyber-contained operations, by definition, only affect the code and content layers of cyberspace.⁹⁰ Prominent models of cyberspace, like for instance the Open Systems Interconnection Model (OSI-model), highlight that cyberspace is not a monolithic structure, but a complex aggregate of different layers.⁹¹ It should be noted that the seven layers of the OSI-model were designed for a different purpose than to legally classify attacks and are sometimes difficult to distinguish in practice. Nevertheless, they are useful with respect to the questions addressed in this article as they allow a disaggregation of cyberspace

⁸⁵ Droege (n 37) 565; Elaine Korzak and James Gow, 'Computer Network Attacks under the Ius Ad Bellum and the Ius in Bello - Distinction, Proportionality, Ambiguity and Attribution', *Routledge Handbook of War, Law and Technology* (Taylor & Francis 2019) 77–78.

⁸⁶ For a similar example, see Robin Geiß and Henning Lahmann, 'Cyber Warfare: Applying the Principle of Distinction in an Interconnected Space' (2012) 45 *Israel Law Review* 381, 385–386.

⁸⁷ Emily Crawford, 'The Principle of Distinction and Remote Warfare' in Jens Ohlin, *Research Handbook on Remote Warfare* (Edward Elgar Publishing 2017) 74.

⁸⁸ NATO Cooperative Cyber Defence Centre of Excellence (n 2) 446.

⁸⁹ Korzak and Gow (n 85) 77–78.

⁹⁰ On the three layers of cyberspace, see Andrew Murray, *The Regulation of Cyberspace: Control in the Online Environment* (Routledge-Cavendish 2006) 54.

⁹¹ Sumit Kumar, Sumit Dalal and Vivek Dixit, 'The OSI Model: Overview on the Seven Layers of Computer Networks' (2014) 2 *International Journal of Computer Science and Information Technology Research* 461.

into a physical layer and other non-physical layers. While physical cyber-infrastructure is as geographically locatable as any other material target, the code and data making up the non-physical layers of cyberspace are harder to conceptualise as targetable entities in a space since they are nothing more than constantly moving electrons.⁹² Although it will sometimes be technically feasible to single out specific sub-systems as targets,⁹³ the general idea of ensuring protection through distinction is based on the assumption of identifiable and separable targets. For this assumption to hold, a significant re-conceptualisation of either the non-physical components of cyberspace or the principle of distinction seems necessary.

Consequently, one might argue that even if the principle of distinction was applicable to operations without physical effects, the interconnected and dual-use nature of cyberspace would make the prohibition on directly attacking civilian objects less relevant. As many virtual components of cyberspace which are also integral for civilian use are likely to qualify as military objectives,⁹⁴ the aspired separation between a predominately military sphere and a predominately civilian sphere becomes increasingly tenuous. Under current LOAC, cyberspace can be conceived of as part of the former sphere based on its pivotal role for the conduct of (both virtual and kinetic) military operations, despite still being overwhelmingly civilian. This 'legal militarisation' of cyberspace runs counter to the overall idea of distinction, namely, to limit attacks to spaces used predominantly by the military.

The concrete legal consequence of this would be that targeting vast parts of cyberspace would be compatible with the prohibition on targeting civilian objects. Civilians would hence not be protected from such operations as such, but only from excessive effects thereof and by the obligation of commanders to take all feasible precautions to prevent such a disproportionate

⁹² Antal Berkes, 'Human Rights Obligations of the Territorial State in the Cyberspace of Areas Outside Its Effective Control' (2019) 52 Israel Law Review 197, 201.

⁹³ In fact, if possible, there is an obligation to do so, according to the TM experts; NATO Cooperative Cyber Defence Centre of Excellence (n 2) 470.

⁹⁴ Terry D Gill, Jelle van Haaster and Mark Roorda, 'Some Legal and Operational Considerations Regarding Remote Warfare: Drones and Cyber Warfare Revisited' in Jens Ohlin, *Research Handbook on Remote Warfare* (Edward Elgar Publishing 2017) 326.

impact. Accordingly, the heavy lifting of protection would have to be done by the principles of proportionality and precaution.⁹⁵

2.1.2 Applying the principles of proportionality and precaution in a complex space

With this conclusion in mind, can the principles of proportionality and precaution be applied to cyberoperations without physical effects in a way that meaningfully limits their impact on civilians? Both principles are crucial to ensure the protection of civilians since they oblige belligerent parties to try to minimise incidental civilian damage while planning and executing operations, and oblige them to refrain from operations that would cause excessive incidental physical harm to civilians. According to the principle of proportionality, as stipulated in Article 51(5)(b) AP1, attacks ‘which may be expected to cause incidental loss of civilian life, injury to civilians, damage to civilian objects, or a combination thereof, which would be excessive in relation to the concrete and direct military advantage anticipated’ are unlawful. The principle of precaution further obliges commanders to ascertain the military nature of their target,⁹⁶ give warnings to civilians if possible,⁹⁷ and to ‘take all feasible precautions in the choice of means and methods of attack with a view to avoiding, and in any event to minimising, incidental loss of civilian life, injury to civilians and damage to civilian objects’.⁹⁸ Furthermore, ‘[w]hen a choice is possible between several military objectives for obtaining a similar military advantage’, the target minimising incidental harm shall be chosen.⁹⁹ Even before planning a specific attack, Article 58 AP1 further sets out general precautionary measures aimed at minimising the incidental impact of military operations on civilians, like the removal of civilians from areas close to military objectives¹⁰⁰ or avoiding the placement of military objectives in the vicinity of densely populated areas.¹⁰¹ One commentator claims that, in principle, these rules apply in the same way to cyberoperations as they do to kinetic operations.¹⁰² This is correct in the sense that the general problems of the principles of proportionality and precaution are

⁹⁵ Geiß and Lahmann (n 86) 398.

⁹⁶ Article 57(2)(a)(i) AP1.

⁹⁷ Article 57(2)(c) AP1.

⁹⁸ Article 57(2)(a)(ii) AP1.

⁹⁹ Article 57(3) AP1.

¹⁰⁰ Article 58(a) AP1.

¹⁰¹ Article 58(b) AP1.

¹⁰² Gill (n 11) 468.

equally pertinent when applied to cyberoperations. However, the following subsection will illustrate how there are three factors aggravating existing challenges¹⁰³ and adding new ones.

First, cyberspace is highly complex. It might not always be straightforward to understand how devices are connected and how they communicate.¹⁰⁴ In many cases, the full impact of an operation will only show after a long time,¹⁰⁵ which makes both the application of Article 57(2)(a) and Article 57(2)(b) AP1 exceptionally difficult. Additionally, the software used by the devices to which a targeted cyberoperation was initially tailored is constantly changing, for instance, due to software updates or minor corrections of coding errors. A malware that might be controllable and spread only within a specific subsystem based on a digital signature (integrated in the software of the devices it is meant to disable) could spin out of control if this signature is removed by a software update, and the malware hence no longer ‘sees’ the difference between devices it is meant to disable and those that should be left unharmed.¹⁰⁶

Second, the direct effects of a cyberoperation on civilian devices and services might be highly heterogeneous because of their differing susceptibility to a specific operation. Coming back to the aforementioned example, malware might only affect users who have already updated their software to the latest version. Consequently, unlike a bomb which severely damages every house it is dropped on, the direct effects of cyberoperations can be highly heterogeneous

¹⁰³ For instance, from a practical perspective, one commentator laments an information asymmetry between commanders and judicial or other bodies scrutinizing their decisions with a view to the principle of proportionality, especially the expected civilian damage and the anticipated military advantage; David Luban, ‘Military Necessity and the Cultures of Military Law’ (2013) 26 *Leiden Journal of International Law* 315, 326–327. From a legal perspective, it is unclear how exactly a military advantage and incidental civilian harm can be compared, and at which exact point the latter becomes excessive; Pilloud and others (n 52) 625.

¹⁰⁴ Roscini (n 62) 95.

¹⁰⁵ Bannelier (n 2) 439.

¹⁰⁶ For a detailed explanation of malware, see Lukasz Olejnik and Tilman Rodenhäuser, ‘Malware: A Selection of Essential Cyber Notions and Concepts’ (*Humanitarian Law & Policy Blog*, 23 May 2019)

<<https://blogs.icrc.org/law-and-policy/2019/05/23/malware-essential-cyber-notions-co-ncepts/>> accessed 16 February 2023.

and depend on the targeted system.¹⁰⁷ This will make it hard to precisely assess the effects of an operation *ex ante*.

Third, assuming for the sake of argument that ‘damage’ is not limited to physical effects or that ‘objects’ includes intangible data, an additional challenge emerges. If non-physical harm is considered, one might wonder how to quantify the damage caused by an operation. Some commentators argue that any loss of functionality should be considered as damage under Article 51(5)(b) AP1 because anything else would make the prediction of damage practically impossible.¹⁰⁸ Yet, this elevates the legal protection of cyber-systems to an end in itself, which it arguably is not. The real harm is not the dysfunctionality or loss of data as such, but the impossibility of using cyberspace or data to enjoy certain freedoms. Consequently, this protection is a means to ensure that civilians can enjoy certain immaterial goods and undertake certain activities.

Given that the actual harm caused to civilians by cyberoperations without physical effects is used as a key argument in favour of moving away from a strict focus on physical damage,¹⁰⁹ it would seem inconsistent to then ignore this actual harm and instead just assume that each loss of functionality or deletion of data will affect civilians in the same way. Yet, if one moves away from this assumption, the assessment of damage would require a ‘translation’ of the inaccessibility of cyber-systems or data into actual harm. This seems problematic as the final harmful effects on civilians’ lives of an operation are likely to differ greatly depending on individual habits, use patterns and the overall dependence on cyber-based services. Imagine an operation disabling a messaging service. If a civilian uses this service to communicate with their family or receive information on the availability of humanitarian aid, they will be greatly affected. Their neighbour, using a different service, might not even become aware of the operation. This example shows that the above-mentioned

¹⁰⁷ Roscini (n 62) 95; Yoram Dinstein and Arne Willy Dahl, ‘Section II: Cyber Operations’ in Yoram Dinstein and Arne Willy Dahl (eds), *Oslo Manual on Select Topics of the Law of Armed Conflict: Rules and Commentary* (Springer International Publishing 2020) 21.

¹⁰⁸ Geiß and Lahmann (n 86) 397.

¹⁰⁹ Lubell (n 36) 275.

‘translation’ might depend on highly individualised features and will hence infinitely complicate an *ex ante* assessment of expected damages.¹¹⁰

This analysis needs to be caveated with respect to one aspect of the principle of precaution: the separation of civilian and military infrastructure (Article 58(b) AP1). Although a complete separation is neither legally required, nor realistic,¹¹¹ the TM lists various preventative measures that states might take, e.g., creating back-ups for essential civilian datasets or promoting antivirus measures to protect civilian systems.¹¹² Such measures seem more feasible than the active precautions addressed above since they do not require an assessment of the precise effects of a specific operation. At the same time, they could prevent the emergence of many harmful effects in the first place by making them technologically impossible or more unlikely, which makes such measures a crucial avenue for protection.

In sum, one might caution that the difficulties of applying the principles of distinction, proportionality and precaution to cyberoperations without physical effects might turn the respective norms, assuming they are applicable, into ‘licensing tools’ (i.e., norms allowing states to legitimise their actions with reference to its lawfulness without actually limiting their discretion in deciding how to conduct hostilities).¹¹³ As explicated above, most of cyberspace could reasonably be construed as a military objective, and anticipating the incidental harm caused by an operation here is especially difficult. Consequently, even for operations which will have a substantial impact on the civilian use of cyberspace, states might reasonably argue that they were targeting a military objective and assess potential collateral effects to the best of their abilities. As one commentator put it, ‘this kind of “normative

¹¹⁰ Note that this is not to say that this problem is equally relevant for all kinds of operations. For instance, an operation affecting transportation services needed to flee from kinetic hostilities will have a relatively homogenous effect on everyone in a certain region. However, there will be operations, like the previously mentioned example, that create highly heterogeneous effects, making an estimation of the expected incidental harm highly speculative.

¹¹¹ Droege (n 37) 567.

¹¹² NATO Cooperative Cyber Defence Centre of Excellence (n 2) 488.

¹¹³ On the licensing effect of LOAC, see Roger Normand and Chris af Jochnick, ‘The Legitimation of Violence: A Critical History of the Laws of War’ (1994) 35 Harvard International Law Journal 49.

indeterminacy” gives belligerents a large margin of appreciation’,¹¹⁴ potentially allowing them to hide from political or moral judgment behind the cloak of plausible legality. Hence, one might wonder how much would actually be gained by making the principles of distinction, proportionality, and precaution applicable – and what might be lost.

III. ‘PLUGIN AVAILABLE’? CIVILIANS MIGHT BENEFIT FROM AN INCREASED FOCUS ON INTERNATIONAL HUMAN RIGHTS LAW

Another assumption that seems to underlie large parts of the above debates is that LOAC is the only body of international law that could provide protection. Arguing that the inapplicability of the principle of distinction causes a gap in protection implies that no other body of law is filling this gap left by LOAC.¹¹⁵ This assumption will be challenged with reference to IHRL.

3.1 Could human rights be the more adequate regime to regulate cyberoperations without physical effects?

Due to its origins as a ‘peacetime law’,¹¹⁶ IHRL is far less focused on violent hostilities than LOAC, and especially the law of targeting. Consequently, the protection it provides is inspired by a fundamentally different logic.¹¹⁷ Instead of prohibiting certain actions, it formulates rights that need to be respected, protected, and fulfilled by states. Hence, IHRL directly protects immaterial values, i.e., rights. This foundation makes IHRL highly relevant to cyberoperations without physical effects since they are liable to precisely impact such immaterial values. Three examples can help to illustrate this point.

First, consider an operation against the head of government of the enemy belligerent party whose private phone calls are intercepted and then

¹¹⁴ Bannelier (n 2) 437.

¹¹⁵ This assumption is further reflected by the focus on LOAC in the academic literature. For instance, the Tallinn Manual discusses LOAC on 189 pages and human rights on 30.

¹¹⁶ Gloria Gaggioli, ‘Remoteness and Human Rights Law’ in Jens Ohlin, *Research Handbook on Remote Warfare* (Edward Elgar Publishing 2017) 154.

¹¹⁷ Cordula Droege, ‘The Interplay between International Humanitarian Law and International Human Rights Law in Situations of Armed Conflict’ (2007) 40 *Israel Law Review* 310, 311.

published online to humiliate her.¹¹⁸ Such an operation is likely to engage the right to private life which has been interpreted by human rights bodies to also encompass the protection of personal data and the privacy of communications.¹¹⁹ In absence of a derogation or a legitimate exception based in law, the publishing of personal data, images or communications would thus be liable to constitute a violation of the right to private life.

Second, a party might decide to block access to all major news portals.¹²⁰ As the internet is increasingly recognised as an important means of exercising the right to freedom of expression,¹²¹ such an operation could constitute a breach of this right. To illustrate: in a case without a nexus to armed conflict, the European Court of Human Rights assessed the legality of a national court order mandating a block on access to a Google service which allowed users to upload content on their websites.¹²² The Court argued that this was the only way to prevent access to certain websites which contained content allegedly insulting the memory of Atatürk.¹²³ The Court considered that the blocked service was 'designed to facilitate the creation and sharing of websites within a group and thus constitutes a means of exercising freedom of expression'.¹²⁴ It ruled that the court order constituted a violation of Article 10 of the European Convention of Human Rights (ECHR).¹²⁵

Third, imagine an operation to delete the online banking information of millions of civilians. As a result, they are unable to access their funds for several weeks and might even permanently lose certain assets. This will certainly have a negative impact on their ability to enjoy their possessions and potentially

¹¹⁸ For a similar operation without a nexus to armed conflict, see Patrick Donahue, 'Germany Suspects Russia-Linked Hackers in Attacks on Politicians' (*Bloomberg*, 29 November 2018)

<<https://www.bloomberg.com/news/articles/2018-11-29/germany-suspects-russia-linked-hackers-in-attacks-on-politicians>> accessed 16 February 2023.

¹¹⁹ *S and Marper v the United Kingdom* [2008] App no 30562/04, 30566/04 (ECtHR, 4 December 2008) [41].

¹²⁰ This resembles certain operations during the Estonia 2007 incident (see above).

¹²¹ See, e.g., *Delfi as v Estonia* [2015] App no 64569/09 (ECtHR, 16 June 2015) [110].

¹²² *Ahmet Yildirim v Turkey* [2012] App no 3111/10 (ECtHR, 18 December 2012) [7–11].

¹²³ *ibid* 51.

¹²⁴ *ibid* 49.

¹²⁵ *ibid* 72.

even amount to expropriation.¹²⁶ Consequently, the right to property might be engaged.

The goal of these examples is not to argue that all these operations would definitely amount to violations of IHRL. They are merely meant to show that IHRL covers a wide range of operations without physical effects, which would not be covered under LOAC if a traditional understanding of violence and objects is maintained since IHRL directly protects the enjoyment of certain rights of freedoms that might be affected by military cyberoperations during armed conflict. More broadly, IHRL might have some generally advantageous features.

Section three of this paper showed that rigid prohibitions might fall short of accounting for the versatile relationship between the direct effect of an operation and its overall consequences for civilian lives. One might argue that this rather rigid and conduct-based nature of the law of armed conflict is the only way to cut through ‘the fog of war’, i.e., the opaqueness emanating from the extreme pressures and dynamic of armed conflict.¹²⁷ However, this argument does not apply in the same way to cyberoperations as it does to traditional operations on a physical battlefield. Cyberoperations require meticulous planning and can usually be conducted far away from the physical battlefield. Hence, it would seem conceivable for commanders to take more nuanced rules into account. IHRL’s rights-based nature allows for a dynamic application of the law as it does not outlaw specific behaviours which are generally likely to cause a certain level of harm, but rather directly protects certain rights, which are then assessed in light of specific state action. As each cyberoperation produces very specific effects, this outcome-based logic might be a better fit than LOAC’s general prohibitions.

Moreover, IHRL’s focus on the individual as a rights-holder and the access to legal remedies (most importantly, court proceedings and potentially reparations) might be an appropriate response to the above-mentioned differential effects of cyberoperations. Given the great difficulties in assessing the precise effects of an operation, even *in the aftermath of an incident*, it might be

¹²⁶ *Sporrong and Lönnroth v Sweden* [1982] App no 7151/75, 7152/75 (ECtHR, 23 September 1982) [61].

¹²⁷ Luban (n 103) 627–628.

sensible to make the individual the relevant unit of analysis and allow for individuals to come forward with complaints regarding how a specific operation has affected them.

Additionally, states would be required to explain why certain operations were necessary to pursue a legitimate aim or were strictly required by the exigencies of a situation.¹²⁸ It is important to emphasise that IHRL would still allow states to derogate from or limit their human rights obligations for security reasons. However, instead of relying on blanket justifications, states would have to show that their derogation or interference was strictly required or necessary to pursue a legitimate aim.¹²⁹ As the burden of proof regarding the lawfulness of an interference or derogation is on the state party,¹³⁰ the application of IHRL might help to ensure a transparent account of how the cyberoperation affected civilians and whether this was strictly necessary. Although states might be reluctant to share the details of why they took a specific measure with reference to the confidential nature of national security, they will have to at least make a generally reasonable case for how an operation was linked to a threat to their security and why the resulting harm to civilians was necessary. While human rights law in general is plagued by an information asymmetry between states and courts in this respect, the mere fact that a case or incident would be discussed in a public forum would make cyberoperations subject to a level of attention and scrutiny, including from civil society actors, which would otherwise be unlikely.

Of course, the details of which institution would deal with claims by individuals, and how exactly it would do so, depends on which IHRL treaty is relied on. While a detailed analysis of specific human rights regimes is beyond the scope of this article, all human rights treaties have some sort of complaint mechanism, albeit differing in their level of enforceability and degree of judicialisation, that could allow for state accountability. The fact that some could

¹²⁸ Article 4(1) ICCPR; Article 15(1) ECHR.

¹²⁹ Marko Milanovic, *Extraterritorial Application of Human Rights Treaties* (Oxford University Press 2011) 254.

¹³⁰ UN Human Rights Committee, 'General Comment No. 29, States of Emergency (Article 4)' (31 August 2001) para 4 <<https://digitallibrary.un.org/record/451555>> accessed 16 February 2023.

do so more effectively than others¹³¹ does not take away from the general observations presented in this article.

In sum, IHRL's focus on individual rights, accountability and transparency could help resolve many of the challenges outlined above with regards to LOAC.

3.2 Problems of applying human rights law

While it was argued that IHRL potentially covers many operations which are likely to escape the reach of LOAC as it stands, human rights' actual protective value depends on one crucial question: is IHRL applicable? For state-conducted cyberoperations, there are two main potential reasons for inapplicability.¹³²

3.2.1 *The relationship between LOAC and IHRL*

The relationship between LOAC and IHRL is a perennial issue. Ever since the International Court of Justice's (ICJ) Advisory Opinion on the *Legality of the Threat or Use of Nuclear Weapons*, there can be no doubt that, in principle, IHRL can apply even in times of armed conflict.¹³³ In its *Wall Advisory Opinion*, the Court further specified that LOAC does not displace IHRL in general. Instead, 'some rights may be exclusively matters of international humanitarian law; others may be exclusively matters of human rights law; yet others may be matters of both these branches of international law'.¹³⁴ Still, there remains

¹³¹ For instance, the ECtHR is certainly exceptional in terms of enforceability since the judgments of the ECtHR are binding upon parties to a specific case (Article 46(1) ECHR) and a failure to abide can entail sanctions (Article 46(5) ECHR).

¹³² In an analysis not limited to state operations, the limited binding effect of IHRL on non-state actors would constitute another such reason. While the legal attribution of cyberoperations conducted by non-state actors to states is beyond the realm of this article, it would need to be explored in an analysis aiming at finding a final verdict on the usefulness of human rights law regarding cyberoperations with a nexus to armed conflict.

¹³³ International Court of Justice, 'Legality of the Threat or Use of Nuclear Weapons, Advisory Opinion' International Court of Justice Reports 1996 [25] <<https://www.icj-cij.org/public/files/case-related/95/095-19960708-ADV-01-00-EN.pdf>> accessed 16 February 2023.

¹³⁴ International Court of Justice, 'Legal Consequences of the Construction of a Wall, Advisory Opinion' International Court of Justice Reports 2004 [106]

ambiguity in how IHRL and LOAC interact in practice. While several ways of conceptualising this interaction are imaginable,¹³⁵ the following discussion will focus on two approaches that are most frequently discussed in the pertinent literature: the *lex specialis* approach and the *belt and suspender* approach. The supporters of the so-called *lex specialis* approach argue that LOAC generally supersedes IHRL as the more specific law in times of armed conflict, unless an issue is not dealt with in the pertinent LOAC treaties.¹³⁶ The alternative *belt and suspender* approach sees LOAC and IHRL on a more equal footing and accords priority to the body of law providing a higher level of protection in a specific situation.¹³⁷ While a full discussion of this issue is beyond the scope of this essay, the specific phenomenon of cyberoperations without physical effects creates an interesting constellation.

Assuming that the law of targeting only applies to acts with reasonably foreseeable physical effects and the law of occupation does not apply,¹³⁸ one

<<https://www.icj-cij.org/public/files/case-related/131/131-20040709-ADV-01-00-EN.pdf>> accessed 16 February 2023.

¹³⁵ While general public international law further knows the *lex superior* (the higher body of law supersedes the lower body of law) and *lex posterior* (the later law repeals the earlier law) approaches, neither are pertinent with a view to this specific question. There is no general rule of superiority between LOAC and IHRL. As per the *lex posterior* approach, it was noted that for ‘overlaps between treaties in different regimes, the question of which of them is later in time cannot be taken to express any intrinsic priority between them’; United Nations General Assembly, ‘Report of the Study Group of the International Law Commission, Finalized by Mr. Martti Koskenniemi’ A/CN.4/L.682 and Add.1 107. Since IHRL and LOAC are different regimes, the principle is not suitable for the questions discussed in this section. In respect of the specific question of how LOAC and IHRL interact, several suggestions have been put forward: parallel applicability, convergence, complementarity, cross-fertilization, and harmonization; Noam Lubell, ‘Parallel Application of International Humanitarian Law and International Human Rights Law: An Examination of the Debate Special Issue: Parallel Applicability of International Humanitarian Law and International Human Rights Law’ (2007) 40 Israel Law Review 648, 654. However, these are rather descriptive or policy-oriented terms than legal doctrines to determine which body of law applies.

¹³⁶ For a general formulation of this reading of the *lex specialis* principle, see Droege (n 117) 338.

¹³⁷ William A Schabas, ‘Lex Specialis - Belt and Suspenders - The Parallel Operation of Human Rights Law and the Law of Armed Conflict, and the Conundrum of Jus Ad Bellum Special Issue: Parallel Applicability of International Humanitarian Law and International Human Rights Law’ (2007) 40 Israel Law Review 592, 598; UN Human Rights Committee (n 130) 3, para 3.

¹³⁸ The role of the law of occupation will be discussed below.

might argue that there is a gap in LOAC that can be filled by IHRL. Potential conflicts between LOAC and IHRL have often been identified with a view to the right to life and detention.¹³⁹ The class of operations considered in this work, *by definition*, does not fall within these areas as they do not foreseeably impact the right to life, physical integrity or free movement.

On the other hand, should the interpretation of the law of targeting evolve to include cyberoperations without physical effects, the rules of LOAC might take precedence under a *lex specialis* approach. If one agrees with the above analysis of the potential shortcomings of LOAC with respect to cyber-contained operations, this dynamic could be seen as another reason to carefully consider the desirability of making the law of targeting applicable to such operations.

However, even if the law of targeting, indeed, were to ‘pull its anchor’ of physical effects, IHRL might still supersede LOAC if it provides more protection under the alternative *belt and suspender* approach.

This essay argues that this could very well be the case.¹⁴⁰ Yet, this would require a rule-by-rule analysis of which rule provides more protection. Given that this depends greatly on the definition of ‘protection’ and the speculated effect of applying a specific norm, there might not always be a clear outcome of such an assessment. Accordingly, this would be a less clear-cut approach than applying IHRL as the more specific law to a situation on which LOAC is silent, which would be possible if the law of targeting remains limited to cyberoperations with physical effects. In the latter case, there would be no issue of conflicting or overlapping norms between LOAC and IHRL and hence no merit in arguments purporting that LOAC supersedes pertinent IHRL guarantees.

¹³⁹ On the right to life, see Droege (n 117), 344-347. On detention, see Lubell (n 135) 657.

¹⁴⁰ A detailed analysis would be required to assess if a similar conclusion can be drawn regarding the law of occupation. For a detailed and critical discussion of applying human rights to cyber-based operations in situations of occupation, see Omar Yousef Shehabi, ‘Emerging Technologies, Digital Privacy, and Data Protection in Military Occupation’ in Russell Buchan and Asaf Lubin (eds), *The Rights to Privacy and Data Protection in Times of Armed Conflict* (NATO CCDCOE Publications 2022).

It should be noted that aside from approaches emphasising the need to determine when LOAC supersedes IHRL and vice versa, the relationship between IHRL and LOAC can also be perceived as one of complementarity.¹⁴¹ This would entail a focus on finding conciliatory interpretations allowing the two bodies of law to ‘influence and reinforce each other mutually’.¹⁴² In fact, it has been argued that most courts have interpreted the *lex specialis* approach to resolve norm conflicts in a way that closely resembles the interpretative principle of systematic integration.¹⁴³ This principle is based on Article 31(3)(c) VCLT, which allows for the consideration of any ‘relevant rule of international law applicable in the relations between the parties’ and is based on the idea that international law should form a coherent set of rules.¹⁴⁴ In the context of this article, the idea of systematic integration is pertinent in resolving some of the interpretative questions and challenges outlined in the two previous sections. To mention a specific example, the term ‘damage’ under LOAC could be interpreted in a way that focuses on the effect on the enjoyment of human rights of a specific operation. A similar proposal could be formulated on the question of which types of data should be protected as civilian objects, namely those which are also protected by human rights guarantees (e.g., the right to private life). In this sense, the challenge of defining and quantifying harm when exclusively focusing on existing norms of LOAC could be mitigated by borrowing from IHRL treaty and case law. However, this approach is premised on the assumption that LOAC and IHRL are generally placed on an equal footing, even during armed conflict. Although certainly relevant in the context of this article, the principle of systematic integration remains a principle of interpretation aimed at reducing situations in which LOAC and IHRL are in conflict with each other, not one of deciding which rule takes precedence for the cases where they still do.

¹⁴¹ Droege (n 117) 340-344.

¹⁴² Droege (n 117) 337.

¹⁴³ Jean D’Aspremont, ‘Articulating International Human Rights and International Humanitarian Law: Conciliatory Interpretation under the Guise of Conflict of Norms-Resolution’ in Malgosia Fitzmaurice and Panos Merkouris (eds), *The Interpretation and Application of the European Convention of Human Rights: Legal and Practical Implications* (Martinus Nijhoff Publishers 2013) 6.

¹⁴⁴ Campbell McLachlan, ‘The Principle of Systematic Integration and Article 31(3)(c) of the Vienna Convention’ (2005) 54 *International and Comparative Law Quarterly* 279, 626; Droege (n 117) 227.

3.2.2 'Physically' strikes back: extraterritorial application of human rights treaties

An equally critical question regarding the applicability of IHRL might be related to its territorial scope. In non-international armed conflict (NIAC), the affected individuals will most likely remain within the territorial jurisdiction of the belligerent state, where the latter is clearly bound by its human rights obligations.¹⁴⁵ On the other hand, in international armed conflict, most operations will most likely affect individuals in the enemy state's territory, who are not necessarily within the acting state's jurisdiction.¹⁴⁶ Consequently, questions pertaining to the extraterritorial applicability of human rights obligations are likely to play a decisive role in determining the protective reach of IHRL.

While positive obligations to protect will bind the belligerent party on whose territory fighting takes place,¹⁴⁷ an attacking belligerent party acting on foreign territory could potentially not be bound by its negative obligations to respect human rights. Since IHRL protections are far more developed in treaty law, the clauses on territorial application in the respective treaties are crucial in this respect.¹⁴⁸ Whereas the International Covenant on Social, Economic and Cultural Rights does not have an explicit territorial limitation, the International Covenant on Civil and Political Rights (ICCPR) and the European Convention on Human Rights (ECHR) apply only to persons within the jurisdiction of the contracting party.¹⁴⁹

Although some states claim that this strictly limits their human rights obligations to their territory,¹⁵⁰ the majority of scholars and the most important

¹⁴⁵ The ECtHR established the assumption that persons in the territory of a Contracting Party are under its jurisdiction, which can only be rebutted in exceptional circumstances. *Ilaşcu and Others v Moldova and Russia* [2004] App no 48787/99 (ECtHR, 8 July 2004) [312].

¹⁴⁶ A more complicated scenario would be an 'internationalised NIAC' in which foreign forces fight on behalf of the government or insurgent forces. In this case, there would be a need to analyse if the intervening foreign forces are bound by their state's human rights obligations.

¹⁴⁷ These obligations have been discussed in detail elsewhere; see, *inter alia*, Berkes (n 92) 227–230.

¹⁴⁸ NATO Cooperative Cyber Defence Centre of Excellence (n 2) 186.

¹⁴⁹ Article 1 ECHR, Article 2(1) ICCPR.

¹⁵⁰ Gaggioli (n 116) 154.

human rights bodies concur that extraterritorial obligations exist.¹⁵¹ However, they remain an exception and are limited to very specific situations. For instance, the case law of the European Court of Human Rights primarily suggests five situations of extraterritorial applicability: all cases of lawful and unlawful occupation,¹⁵² specific acts of (physical) violence involving an element of proximity,¹⁵³ actions taken by diplomatic and consular agents abroad, exercises of extra-territorial public powers at the invitation or with the acquiescence of a foreign state and situations in which a state has *effective control* over a person or territory of a foreign state.¹⁵⁴ The final situation can be understood as *physical* control over territory or a person. The Court had stated earlier that the ECHR does not apply to airborne operations with lethal effects outside the European legal space. In fact, it explicitly rejected a ‘cause-and-effect’ notion of territorial applicability similar to that of the Geneva Conventions.¹⁵⁵

If applying an ‘effective control’ standard or similar theories, it is unlikely that a cyberoperation would trigger extraterritorial human rights obligations unless it affects persons detained by the belligerent party or are situated in occupied territories. However, since cyberoperations do not require physical control over territory or persons to produce significant effects on the enjoyment of rights by people in this territory,¹⁵⁶ human rights treaties might remain inapplicable in situations where cyberoperations produce significantly harmful effects in territories which are not under effective control of the acting belligerent party.

¹⁵¹ *ibid* 155.

¹⁵² *Loizidou v Turkey* [1996] App no 15318/89 (ECtHR, 18 December 1996) [52].

¹⁵³ *Georgia v Russia (ii)* [2021] App no 38263/08 (ECtHR, 21 January 2021) [130–131]; *Carter v Russia* [2021] App no 20914/07 (ECtHR, 21 September 2021) [129].

¹⁵⁴ *Al-Skeini and Others v the United Kingdom* [2011] App no 55721/07 (ECtHR, 7 July 2011) [74]. The UN Human Rights Committee used the formulation ‘power or effective control’, which raises a similar issue. UN Human Rights Committee, ‘General Comment No. 31, The Nature of the General Legal Obligation Imposed on States Parties to the Covenant’ (29 March 2004) para 10

<<https://digitallibrary.un.org/record/533996?ln=en>> accessed 16 February 2023.

¹⁵⁵ *Banković and Others v Belgium* [2001] App no 52207/99 (ECtHR, 12 December 2001) [75–80].

¹⁵⁶ On the spatial model, see Milanovic (n 129) 127–173. On the personal model, see *ibid* 173–209.

The majority of the TM experts equally considered physical control over either territory or individuals as a necessary premise for human rights to apply extraterritorially. To support this view, they cited the reluctance of many states to apply IHRL to their intelligence operations targeting foreigners abroad.¹⁵⁷ However, it should be noted that the US submitted their intelligence programme to human rights obligations at least as a matter of policy.¹⁵⁸ A similar move had also been proposed by several states in a draft UN General Assembly resolution.¹⁵⁹ Although these developments fall short of establishing extraterritorial obligations for states under the law as it stands, they might still be interpreted as evidence for a general willingness to reconsider the territorial scope of international human rights law with respect to cyber-based behaviour. Moreover, it might be noteworthy that states increasingly tend to extend their jurisdiction to foreigners when regulating online platforms.¹⁶⁰ While there is a great difference between general jurisdiction and jurisdiction as the delimitator of the territorial scope of human rights obligations, this development might nevertheless signify a general trend towards a less territorial understanding of jurisdiction in the context of cyber-activities.

Such an evolution would be in line with the *sui generis* nature of cyberspace. While the infrastructure making up the physical layer of cyberspace can be assumed to be under the exclusive control of one state,¹⁶¹ the code and content layer of cyberspace are hard to fit into the Westphalian model of

¹⁵⁷ NATO Cooperative Cyber Defence Centre of Excellence (n 2) 185.

¹⁵⁸ David P Fidler, 'Cyberspace and Human Rights' in Nicholas Tsagourias and Russell Buchan (eds), *Research Handbook on International Law and Cyberspace* (Edward Elgar Publishing 2021) 147.

¹⁵⁹ *ibid* 145.

¹⁶⁰ The EU General Data Protection Regulation applies to the processing of personal data by EU citizens regardless of where the processing takes place; Regulation (EU) 2016/679 of the European Parliament And of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [2016] OJ L119/1, Article 3(1). The EU Digital Services Act will apply to providers who 'offer services in the Union'. Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC [2020] COM(2020) 825 final, Article 2(d).

¹⁶¹ Berkes (n 92) 207.

territorially grounded sovereignty.¹⁶² While servers and cables are located in one place, computer commands and online content are nothing more than electrical signals travelling within milliseconds from one device or server to another. Therefore, it will be very difficult, if not impossible, to use their physical location as a way of deciding which laws apply.¹⁶³ Although it would be premature to derive any consequences for the applicability of human rights *lex lata* from this observation, it might nevertheless serve as a strong argument in favour of reconsidering existing concepts on the extraterritorial application of IHRL. In fact, a minority of the TM experts argued that ‘so long as the exercise or enjoyment of a human right in question by the individual concerned is within the power or effective control of a State’, human rights law should apply.¹⁶⁴ Furthermore, it should be noted that the ECtHR recently held that ‘[t]he extent of the State’s obligations under Article 1 of the Convention is to secure to [an] individual the Convention rights and freedoms that are relevant to his or her situation’.¹⁶⁵ While this holding did not relate to cyber-based conduct, it could nevertheless be seen as a sign that the ECtHR is willing to move towards a more flexible, effects-based approach to extraterritorial applicability.

This is not to say that such an approach would be entirely unproblematic. The ECtHR in *Banković* expressed the fear that a ‘cause-and-effect’ notion of extraterritorial applicability would fragment the applicability regime of the Convention into several pieces that are specifically tailored to the particular circumstances of an extraterritorial act, hence negatively impacting the integrity of the Convention.¹⁶⁶ Furthermore, there might be a need to distinguish between positive and negative obligations, with

¹⁶² Dan Efrony and Yuval Shany, ‘A Rule Book on the Shelf? Tallinn Manual 2.0 on Cyber Operations and Subsequent State Practice’ (2018) 112 *American Journal of International Law* 583, 653.

¹⁶³ David R Johnson and David Post, ‘Law and Borders: The Rise of Law in Cyberspace’ (1996) 48 *Stanford Law Review* 1367, 1370–1371.

¹⁶⁴ NATO Cooperative Cyber Defence Centre of Excellence (n 2) 185.

¹⁶⁵ *Ukraine and the Netherlands v Russia (dec)* [2022] App no 8019/16, 43800/14, 28525/20 (ECtHR, 30 November 2022) [571].

¹⁶⁶ *Banković and Others v Belgium* (n 155) para 75. Note, however, that this conclusion has recently been explicitly overruled by the Court in its admissibility decision on applications arising from the conflict between Russia and Ukraine; *Ukraine and the Netherlands v. Russia (dec.)* (n 165) para 571.

only the latter applying extraterritorially based on the effect of a state act.¹⁶⁷ If any act of a state having some effect on an individual in another state triggered obligations to protect and to fulfil the human rights of said individual, this could lead to conflicts with state sovereignty and overwhelm individual states.¹⁶⁸ Additionally, if any type of extraterritorial effect were to trigger the applicability of human rights treaties, the respective courts and mechanisms might quickly be overwhelmed by a flood of claims.¹⁶⁹

Despite these challenges, the above considerations regarding the desirability of applying human rights law to military cyberoperations lead this author to conclude that debates on military cyberoperations should not only focus on updating LOAC to present-day warfare, but also consider the need to adapt human rights law to the unique nature of cyberspace. As IHRL protects immaterial goods, it is apt to close gaps in legal protection left by LOAC's focus on physicality. However, it can only do so if it is applicable. This is currently not the case when a state uses cyberspace to cause harm in the territory of an enemy belligerent state. This author believes that there is a need to extend the applicability of IHRL to such conduct so as to ensure the full enjoyment of human rights.

IV. THE WAY AHEAD: A PLEA TO BROADEN THE DEBATE ON MILITARY CYBEROPERATIONS WITHOUT PHYSICAL EFFECTS

Sections two and three explicated why the most prevalent rules of the law of targeting are not easily transferable to cyber-contained operations. Section four argued that IHRL's rights-based approach might be more apt to the specific challenges of such operations but conceded that its utility in IAC is strongly circumscribed by its limited territorial applicability.

Hence, this author concurs with many scholars that a gap in protection exists. However, the above discussions suggest that it is not a lack of physical effects alone which causes a gap in protection. Instead, combining the results of

¹⁶⁷ Milanovic (n 129) 209–219; Marko Milanovic and Michael N Schmitt, 'Cyber Attacks and Cyber (Mis)Information Operations during a Pandemic COVID-19 Special Edition: Part IV: Comparative and International Law Perspectives' (2020) 11 *Journal of National Security Law and Policy* 247, 263.

¹⁶⁸ Milanovic (n 129) 219.

¹⁶⁹ *ibid* 221.

the preceding sections, the root of the problem can be traced back to an underlying assumption shared by both LOAC and IHRL: immaterial values/rights can only be *directly and significantly* restricted if a party has physical control over persons or territory.

4.1 The real problem: The relation between virtual and physical space

It would be an oversimplification to suggest that the LOAC as a whole is ‘physically anchored’. In fact, non-material values are protected by several provisions: Article 75(2)(b) AP1 protects against outrages upon personal dignity, Article 25 Geneva Convention IV (GCIV) protects communications with family members, Article 27 GCIV guarantees protected individuals ‘respect for their persons, their honour, their family rights, their religious convictions and practices, and their manners and customs’. All these provisions have two commonalities: they directly protect human dignity¹⁷⁰ and they all only apply once a person is ‘in the hands’ or ‘in the power’ of a belligerent party.¹⁷¹ This shows that LOAC does in fact provide direct protection to non-material goods, but it limits such protection to situations where the belligerent party has physical control over a person or the territory this person finds themselves in.¹⁷²

Against this background, the discussions on Article 54 AP1 shed new light on the function of the term ‘attack’.¹⁷³ Initially, the ICRC had proposed two quasi-identical provisions on the protection of objects indispensable to the survival of the civilian population. One was placed in the section on the general protection of the civilian population against the effects of hostilities, the other one was placed in the section on the treatment of persons in the power of a party to the conflict. The former (draft Article 48) stated that it was ‘forbidden to *attack* or destroy’ certain indispensable objects,¹⁷⁴ whereas the latter (draft

¹⁷⁰ Jean Pictet and others, *Commentary on the Geneva Conventions of 12 August 1949*, vol 4 (ICRC 1958) 200.

¹⁷¹ See Article 4(1) GCIV and Article 75(1) AP1.

¹⁷² Pilloud and others (n 52) 47; *ibid* 866.

¹⁷³ Given the persistent unclarity and potentially absurd conclusions regarding Article 49(1) AP1 outlined in section 2, a recourse to the preparatory work under Article 32 VCLT seems warranted.

¹⁷⁴ One might be confused by the mentioning of both ‘to attack’ and ‘to destroy’. However, this does not imply that to attack and to destroy were considered mutually exclusive concepts. Diplomatic Conference on the Reaffirmation and Development of

Article 66) made it unlawful to ‘destroy, *render useless, or remove* objects indispensable to the survival of the civilian population’, which resembles the functionality approach proposed by some scholars with respect to cyber-contained operations. The difference between the two draft provisions was explained by an ICRC representative in the following terms: ‘The words “to attack” were not included in draft Article 66 because *one did not attack what was in one’s possession*’.¹⁷⁵ This shows that the drafters assumed that attacks are only mounted by the party which is not in control of the objects or persons it is attacking. Any protection against attacks would hence only include those goods and values which were directly targetable while a party is not in control of the objects and persons targeted. In 1977, when the Additional Protocols to the Geneva Conventions were adopted, the private and family life of an individual (including the possibility to communicate), their personal dignity and other non-material goods, in absence of physical control over a person or territory, were only indirectly targetable via the physical destruction of certain objects required to give effect to them. Outlawing acts of violence targeting or excessively affecting civilian objects hence automatically also covered such non-material goods and liberties. Yet, this protection is always translated into the only way in which a belligerent party could impact civilian life in absence of territorial control: violence. As immaterial aspects of civilian life move online and become directly targetable through cyber-means, a mere extension of the scope of the law of targeting might not be sufficient since the actual harm caused by cyberoperations without physical effects simply cannot be translated into the language of violence and objects. Even if the existing rules were applicable, their protective value might be greatly diminished by the challenges faced when trying to apply the rules that were meant to protect material objects to the protection of immaterial objects and values more broadly. In this way, their protective value might be ‘lost in translation’.

In short, the author avers that there is a need to interpret the focus on physicality in the law of targeting, not only as a result of the means and methods of warfare known at the time when the Geneva Conventions and their Protocols were drafted. Instead, it should also be seen as a consequence of the drafters’ implicit assumptions about how civilian life can possibly be affected by a

International Humanitarian Law Applicable in Armed Conflicts Geneva (n 48) 395, para 36.

¹⁷⁵ *ibid*, para 36.

belligerent party in the absence of physical control. Consequently, not only the notion of ‘violence’, but also the concept of having ‘control over a person’ need to be taken into account when trying to ensure that civilians are as well protected from offline as from online harms.

This article has explored whether IHRL might close gaps in the legal protection provided by LOAC arising from the previously discussed features. It is concluded that its capacity to do so is limited by a similar assumption. IHRL is based on the idea that individuals have certain rights which can potentially be limited by the state since the latter has the monopoly of force with regards to the individuals within its territory. As a result of this power, a state has positive and negative obligations to ensure the enjoyment of these rights. While IHRL is conceptually more apt to regulate the harm caused by cyberoperations without physical effects since it protects immaterial values, a corollary of IHRL’s focus on a state’s ability to limit or facilitate the enjoyment of human rights is that its applicability is limited to situations where a state actually is ‘in power’.¹⁷⁶ The latter notion has been interpreted by many human rights bodies as physical control over a person or territory. This author hence concludes that protection could be provided by IHRL, if it did not rely on the same fundamental notion as LOAC does: to be able to affect a person’s rights and freedoms, a state must have physical control over that person.

Cyberspace, however, allows parties to have remote control over the virtual space in which important aspects of civilian lives (communications, political participation, access to property and essential services, etc.) manifest without any physical control. In the view of the author, this is the actual novel feature of cyberspace causing a gap in legal protection from cyberoperations without physical effects. To legally and practically close the existing gap in protection, there would hence be a need to combine IHRL’s direct and comprehensive protection of immaterial goods and the law of targeting’s recognition that states also have to respect the civilian population of areas over which they have no effective control during armed conflict.

¹⁷⁶ Droege (n 117) 335.

4.2 Possibility of an integrated approach?

Consequently, it seems reasonable to look for ways to ‘fold IHRL into LOAC’. Two potential avenues for integrating IHRL into LOAC in certain respects will be briefly discussed here: legal weapon reviews and the Martens Clause.

4.2.1 Legal reviews of cyberweapons

‘In the study, development, acquisition or adoption of a new weapon, means or method of warfare’, parties to AP1 are ‘under an obligation to determine whether its employment would, in some or all circumstances, be prohibited by this Protocol *or by any other rule of international law applicable to the High Contracting Party*’.¹⁷⁷ The wording of Article 36 AP1 suggests that this review is to be undertaken by states with a view to their weapons, means or methods of warfare’s compatibility with *all* their obligations under international law. The review is consequently not limited to LOAC but could also include IHRL.¹⁷⁸ However, there are three shortcomings to this approach.

First, Article 36 AP1 is notoriously weakly implemented. Only a few states have publicly stated that they conduct regular reviews of new weapons they develop or acquire¹⁷⁹ and there is no international body or forum to oversee the review process.¹⁸⁰ Second, states might argue that if they do not cause any physical effects, cyber-capabilities do not qualify as ‘weapon, means or method of warfare’ and hence fall outside the scope of Article 36 AP1.¹⁸¹ Third, even when a cyber-capability is not unlawful under IHRL by nature, it could nevertheless be used in a manner incompatible with IHRL. As cyber-capabilities usually exploit specific weaknesses in the enemy’s system, nature and use of

¹⁷⁷ Article 36 AP1 (emphasis added).

¹⁷⁸ Kathleen Lawand, ‘Reviewing the Legality of New Weapons, Means and Methods of Warfare’ (2006) 88 International Review of the Red Cross 925, 929.

¹⁷⁹ William H Boothby, *Weapons and the Law of Armed Conflict* (Oxford University Press 2016) at 19.5

<<https://opil.ouplaw.com/view/10.1093/law/9780198728504.001.0001/law-9780198728504>> accessed 7 March 2020.

¹⁸⁰ Pilloud and others (n 52) 426.

¹⁸¹ Duncan Blake and Joseph S Imburgia, ‘Bloodless Weapons - The Need to Conduct Legal Reviews of Certain Capabilities and the Implications of Defining Them as Weapons’ (2010) 66 Air Force Law Review 157, 168.

cyber-capabilities might be identical in some cases. For instance, the *Stuxnet* code was specifically designed for one specific target, namely the turbines in the targeted nuclear facility. On the other hand, general techniques (e.g., DoS attacks overwhelming servers with requests), could be considered one class of weapons which is not problematic by nature, but only if used in a certain manner. Imagine a DDoS operation that is strictly limited to military devices and has no spill-over effects on the lives of civilians. This operation would most likely not infringe any human rights. However, the very same technology could be used to render all cash machines and online banking tools dysfunctional, hence affecting the right to property. The general method of overwhelming a service or device with requests and hence causing it to be inaccessible might thus make it through a legal review and could still be used in a way that will affect the enjoyment of human rights.

In addition to these challenges, cyber capabilities often need to be constantly modified and updated to respond to changes in the target system. Conducting a new review after every change, however minimal, might be infeasible.¹⁸² However, despite the continuous, iterative nature of the review process,¹⁸³ such a narrow interpretation of the term *new* would hardly be reconcilable with the purpose of Art 36 API, which is to prevent the use of certain weapons having unlawful *effects*.¹⁸⁴ Therefore, it seems reasonable to only conduct a new review if the modification in question substantially alters the weapon's intended effects or operational performance.¹⁸⁵ Nevertheless, the constantly evolving nature of cyber capabilities is a general complicating factor in applying Article 36 API.

Consequently, emphasising states' obligation to consider human rights guarantees when reviewing new weapons, means and methods of warfare might

¹⁸² Gary D Brown and Andrew O Metcalf, 'Easier Said than Done: Legal Reviews of Cyber Weapons' (2014) 7 Journal of National Security Law and Policy 115, 128.

¹⁸³ Marcos Kotlik, 'Reviewing Legal Weapons Reviews: Is It Possible to Verify Compliance?' (*EJIL: Talk!*, 19 March 2020) <<https://www.ejiltalk.org/reviewing-legal-weapons-reviews-is-it-possible-to-verify-compliance/>> accessed 20 March 2020; Justin McClelland, 'The Review of Weapons in Accordance with Article 36 of Additional Protocol I' (2003) 85 International Review of the Red Cross 397, 413.

¹⁸⁴ This interpretation is based on the second part of Article 31(1) VCLT.

¹⁸⁵ McClelland (n 183) 404; NATO Cooperative Cyber Defence Centre of Excellence (n 2) 466.

be useful under very specific circumstances, namely if a state actually has a weapons review procedure and has developed or acquired a cyber-capability which would in all foreseeable use-cases be incompatible with a state's human rights obligations. Such a situation is not entirely unthinkable,¹⁸⁶ but it will most likely be the exception rather than the rule. Hence, given the generally weak enforcement of Article 36 AP1 and the aforementioned cyber-specific challenges, the utility of this approach in ensuring comprehensive protection for civilians will most likely be very limited.

4.2.2 Martens Clause

The Martens Clause (Article 1(2) AP1) was included in the text of AP1 for situations which might arise after the conclusion of the treaties and without being foreseen by the drafters who could not possibly predict all future technological developments.¹⁸⁷ The provision states that '[i]n cases not covered by this Protocol or by other international agreements, civilians and combatants remain under the protection and authority of the principles of international law derived from established custom, from the principles of humanity and from the dictates of public conscience'.¹⁸⁸ In light of the fact that the above-identified gap of protection was caused precisely by such a technological development, the Martens Clause seems to be a likely candidate for filling this gap.

Its reference to 'established custom' could be interpreted to include customary IHRL.¹⁸⁹ Nevertheless, to argue that the Martens Clause would allow to simply 'import' the substance of IHRL into LOAC without heeding the former's territorial limits would arguably meet fierce resistance from states

¹⁸⁶ Note that several states (e.g., Belgium, Germany, The Netherlands, New Zealand, Norway, Sweden, Switzerland, the UK and the US) have publicly declared that they conduct legal reviews of new weapons; Vincent Boulanin and Maaïke Verbruggen, 'SIPRI Compendium on Article 36 Reviews' (SIPRI 2017) 2 <<https://www.sipri.org/publications/2017/sipri-background-papers/sipri-compendium-article-36-reviews>> accessed 16 February 2023. Should these states develop or acquire a capability that is exclusively designed to, e.g., disable all online-based ways of receiving information and expressing one's opinions, this might be considered incompatible with human rights obligations.

¹⁸⁷ Pilloud and others (n 52) 39, para 55.

¹⁸⁸ Article 1(2) AP1.

¹⁸⁹ Rupert Ticehurst, 'The Martens Clause and the Laws of Armed Conflict' (1997) 37 *International Review of the Red Cross* (1961 - 1997) 125, 126.

which are generally reluctant to accept extraterritorial obligations. Indeed, since international law is based on, *inter alia*, the principle of state consent, one might reasonably argue that the Martens Clause cannot be used to circumvent applicability provisions in IHRL treaties states explicitly agreed to. Given the general disagreement about the Clause's precise content and its legal effect,¹⁹⁰ it seems likely that states and commentators would not easily accept such a far-reaching interpretation of Article 1(2) AP1. Furthermore, some of the human rights protections relevant in the context of cyberoperations (e.g., the right to property) might not be widely accepted as customary law. Hence, they would not form part of 'established custom', but at most 'the laws of humanity'. However, the inherent vagueness of this category further decreases the strength of the argument that the Martens Clause allows to apply IHRL norms to cyberoperations with a nexus to armed conflict.

Yet, the Martens Clause could be useful to encourage and legally ground interpretations of existing LOAC which avoid results that are incompatible with IHRL guarantees. This constitutes a rather soft, IHRL-driven bending of LOAC instead of a robust IHRL implant since IHRL would not be directly applicable, but merely used to interpret existing LOAC. Such an approach could nevertheless be helpful to counteract an unfortunate neglect of IHRL when discussing the regulation of military cyberoperations without physical effects. In sum, the Martens Clause might be helpful in supporting the argument that LOAC should be interpreted in a way that would establish legal protection for civilians from all severely harmful cyberoperations, regardless of the physicality of their effects. However, to argue that IHRL protections apply in any situation of armed conflict because of Article 1(2) AP1's reference to established custom or the laws of humanity would go too far in the view of this author. Consequently, the Martens Clause cannot fill existing gaps in the legal protection of civilians from cyberoperations without physical effects.

In sum, neither of the two approaches discussed here seem to offer truly promising avenues of combining the substantive protections of IHRL with the territorial reach of LOAC. As has been shown, taking IHRL guarantees into account during weapons reviews is unlikely to have a significant positive impact

¹⁹⁰ Schmitt, 'Wired Warfare 3.0' (n 53) 354; A Cassese, 'The Martens Clause: Half a Loaf or Simply Pie in the Sky?' (2000) 11 European Journal of International Law 187, 188; Ticehurst (n 189) 126.

on the protection of civilians from cyberoperations without physical effects due to the general nature of Article 36 AP1 reviews and its weak implementation. Consequently, only very few situations, if any, in which a capability would be discarded by states as the result of an IHRL-based weapons review are conceivable. The Martens Clause offers a broader way of ‘folding IHRL into LOAC’. However, this author does not believe that the Clause can be stretched so far as to justify the circumvention of explicit provisions in IHRL treaties limiting their territorial applicability. Hence, Article 1(2) AP1 might be a useful stanchion for argumentative construction when militating in favour of IHRL-informed interpretations of existing LOAC, but it will most likely not be more than that. Consequently, neither Article 36 AP1, nor Article 1(2) AP1 can resolve the dilemma set out at the very beginning of this subsection.

V. CONCLUSION

This paper has critically evaluated a common assumption of the literature regarding military cyberoperations without any reasonably foreseeable physical effects with a nexus to armed conflict. The main argument put forward in this respect is that even if the ‘physical anchoring’ of the law of targeting is undone, substantial conceptual and practical challenges remain. Most notably, the protective potential of the principle of distinction is decreased by the interconnected and complex nature of cyberspace, which might make large segments of the internet and its non-physical infrastructure legitimate targets and severely complicate proportionality assessments. Consequently, there is a risk that if the principles of distinction, proportionality and precaution apply, they will serve as permissive ‘licensing tools’, rather than provide effective protection to civilians.

It was submitted that IHRL’s broader substantive scope as well as its focus on individual rights, transparency and accountability seem to be a better fit for the idiosyncratic and complex effects of cyber-contained operations. Yet, its practical utility is greatly limited in IACs: as long as extraterritorial obligations only arise if a state physically controls a person or territory abroad, most cyberoperations with a nexus to armed conflict will escape the reach of human rights obligations. This paper therefore argued that an effects-based approach to the extraterritorial application of human rights treaties should be discussed more systematically in this context.

Connecting both observations, it was purported that the true gap in protection lies not only in the law of targeting's focus on violent effects. The *travaux préparatoires* to AP1 suggest that the law of targeting was only considered relevant to operations against persons and objects outside the physical control of the acting belligerent party. In this sense, it is not only physically, but also territorially anchored. In absence of physical control over territory, non-material values and goods could only be affected through the destruction of certain objects, which is why protection therefrom effectively covered both material and non-material goods and values. The rules of the law of targeting hence did not need to be tailored to the protection of immaterial goods like privacy, immaterial property or free expression. How has this changed? Due to the increased societal importance of cyberspace, a person's dignity, family life and democratic freedoms are now *directly* targetable even without physical control over territory. Hence, civilians are at risk of falling into a gap between the law of targeting, which seems inadequate even if applicable, and IHRL or even the law of occupation which might be more adequate but inapplicable in absence of physical control.

In the end, there is no panacea to the issues raised by military cyberoperations without physical effects with a nexus to armed conflict. Some have suggested that new norms are needed to satisfyingly regulate military cyberoperations.¹⁹¹ However, there seems to be little motivation on the side of states to negotiate a new treaty or similar instrument.¹⁹² Assuming that existing rules of international law will remain the relevant framework for the regulation of military cyberoperations for the foreseeable future, this article suggests that there are theoretical and practical reasons to 'fight on all fronts', including IHRL, instead of focusing predominantly on LOAC. Importantly, the observations and arguments put forward in this paper are far from sufficient to

¹⁹¹ Brad Smith, 'The Need for a Digital Geneva Convention' (*Microsoft On the Issues*, 14 February 2017)

<<https://blogs.microsoft.com/on-the-issues/2017/02/14/need-digital-geneva-convention/>> accessed 16 February 2023; Louise Arimatsu, 'A Treaty for Governing Cyber-Weapons: Potential Benefits and Practical Limitations', *2012 4th International Conference on Cyber Conflict (CYCON 2012)* (2012) 91.

¹⁹² Michael N Schmitt and Sean Watts, 'The Decline of International Humanitarian Law Opinio Juris and the Law of Cyber Warfare' (2015) 50 *Texas International Law Journal* 189, 222.

conclude that IHRL will provide better protection than LOAC in all conceivable situations. Instead, the far more limited proposition of this paper is that the role of IHRL regarding military cyberoperations should be analysed more thoroughly and taken into account when discussing how LOAC should respond to the increased use of cyber-based means during armed conflict. More specifically, further research would be needed to assess if and to what extent the utility of IHRL is limited by its focus on state actors when non-state groups are among those mounting cyberoperations. Furthermore, a more detailed investigation of the possibility to derogate from human rights obligations during armed conflict and to impose lawful limitations requires scholarly attention.

With a view to Ukraine, the arguments presented in this article suggest that there is some merit in not only focusing on cyberoperations as war crimes, but also as human rights violations. While the European Court of Human Rights might no longer be available as a forum,¹⁹³ an increased focus on human rights by civil society and academia might allow for the condemnation of operations which currently do not qualify as war crimes, based on their impact on the enjoyment of human rights. This might also trigger a debate on how to interpret jurisdictional clauses in human rights treaties with a view to cyber-based conduct, which is what might currently prevent the classification of Russian cyberoperations with effects in Ukraine as actual human rights violations.

More generally, as the ongoing conflict in Ukraine shows, powerful states are taking actions which might turn cyberspace into ‘just another domain of warfare’. Legal scholars should critically evaluate to what extent they might acquiesce to such a transformation when predominantly focusing on the rules on targeting while a body of law is available that more directly protects what is at stake, *inter alia*, for Ukrainians affected by cyberoperations connected to the Russian invasion: the enjoyment of a life in freedom and dignity, both online and offline.

¹⁹³ Council of Europe, ‘Russia Ceases to Be Party to the European Convention on Human Rights’ (*Council of Europe Newsroom* 16 September 2022) <<https://www.coe.int/en/web/portal/-/russia-ceases-to-be-party-to-the-european-convention-on-human-rights>> accessed 16 February 2023. There would still be a possibility to bring complaints under the individual complaint mechanism established by the First Additional Protocol of the ICCPR.

SPRING ISSUE



2023

Decentralisation and Recentralisation: An Institutional Analysis of EU Competition Law and the Digital Markets Act

Ruairí McIntyre*

ABSTRACT

The recently enacted Digital Markets Act Regulation (DMA) has revitalised discussion regarding the engagement of national actors in the enforcement of EU competition law. Reversing a trend of consistent decentralisation since the ‘legal and cultural revolution’ of modernisation in 2004, the new Regulation concentrates the authority to apply specialised competition rules within the exclusive competence of the European Commission. Though somewhat necessitated by the peculiar conditions of digital markets, this Regulation reveals a familiar reluctance within the Commission to devolve responsibility to national competition authorities (NCAs), citing a ‘risk of divergence’ and fear of ‘fragmentation’ in defence of centralised enforcement. Little attention has been afforded to a further institutional divergence underlying the motivation and momentum behind decentralisation; a divergence occurring not between national authorities, but between central EU institutions. This article explores institutional tension between the Court of Justice of the European Union (CJEU) and the European Commission in the field of EU competition law. It seeks to demonstrate that the essential structural, economic, and political characteristics of EU competition law have been wrought by divergence between the jurisprudence of the former and policies and practice of the latter. Four phases of the development of EU competition law are identified, and the emergence of a fifth is suggested: ‘Centralisation’, ‘Decentralisation’, ‘Modernisation’, ‘Fragmentation’, and (potentially) ‘Recentralisation’.

* BCL, University College Cork (National University of Ireland). UCC Quercus Scholar 2021/2022. I would like to thank the editorial board of the LSE Law Review for their patience and assistance. Any mistakes are my own.

INTRODUCTION

The recent enactment of the Digital Markets Act (DMA)¹ has revitalised discussion regarding the engagement of national actors in the enforcement of EU competition law. Reversing a trend of consistent decentralisation since the ‘legal and cultural revolution’ of modernisation in 2004,² the new Regulation concentrates the authority to apply specialised competition regulations within the exclusive competence of the European Commission, resurrecting a centralised executive monopoly that harkens back to a former period in the history of EU competition regulation. Though somewhat necessitated by the peculiar conditions of digital markets, this Regulation reveals a familiar antipathy within the Commission towards the devolution of responsibility to national competition authorities (NCAs), citing a ‘risk of divergence’ and fear of ‘fragmentation’ in defence of centralised enforcement.³

Academic discourse in the field of competition law has sought to validate and substantiate this cardinal concern of the Commission. Significant research has been devoted to examining procedural and substantive divergence arising from the disparate application and enforcement of EU competition provisions by national institutions, studying the different attitudes, philosophies and constitutions of various national competition authorities.⁴ However, little attention has been afforded to a further institutional divergence underlying the motivation and momentum behind decentralisation; a divergence occurring not between national authorities, but between central EU institutions. This article explores the tension between the Court of Justice of the European Union (CJEU) and the European Commission in the field of EU competition law,

¹ Council Regulation (EC) 2022/1925 on contestable and fair markets in the digital sector (Digital Markets Act) [2022] OJ L 265/1.

² Claus-Dieter Ehlermann, ‘The Modernization of EC Antitrust Policy: A Legal and Cultural Revolution’ (2000) 37 *Common Market Law Review* 537, 537.

³ See the explanatory memorandum to the DMA’s Proposal Document: Proposal for a Regulation of the European Parliament and of the Council on contestable and fair markets in the digital sector (Digital Markets Act), COM (2020) 842 final (Brussels, 2020).

⁴ See, for example, Rebecca Walker, ‘Procedural Divergence as a Barrier to Effective EU Competition Law Enforcement in Digital Commerce’ (2021) 42(8) *European Competition Law Review* 468.

examining the divergence between the jurisprudence of the former and the policies of the latter. It seeks to explain this institutional tension and assess the decentralised regime of competition law wrought thereby.

It is hoped that this examination of institutional discordance shall inform future analysis of decentralisation within EU competition law by providing an enriched understanding of an institutional dynamic pivotal to the definition and evolution of EU competition law. Whilst a wealth of academic literature thoroughly examines decentralisation as a *cause* of institutional divergence, it rarely assesses decentralisation as a *consequence* thereof.⁵ This article seeks to provide such an assessment through a detailed historical account of the development of the EU competition law regime, sketching the history of decentralisation in light of this conflict.

Following this introduction, the article identifies four discrete ‘phases’ in the development of European competition policy, and suggests the emergence of a fifth. Section II, examining ‘centralisation’, illustrates the collaboration of the Court and Commission in establishing an effective and comprehensive competition law regime. Section III, entitled ‘decentralisation’, analyses the tentative model of decentralisation prior to 1999, constructed around private enforcement. Section IV examines ‘modernisation’ and the Commission’s redefinition of decentralised competition enforcement. Section V assesses the resultant ‘fragmentation,’ and the contribution of the Court of Justice of the European Union (CJEU) thereto. Section VI discusses the apparent emergence of a new phase of ‘recentralisation’ in legislative policy. Section VII concludes. In examining decentralisation, a phenomenon necessarily involving the engagement of national actors, this article will also explore further institutional relationships influenced by, and directly connected to, this central dynamic. In analysing the Court’s promotion of private enforcement of competition law, Section III discusses the relationship between the CJEU and national courts. In assessing the Commission’s promotion of decentralised public enforcement to counteract the effects of private enforcement, Section IV

⁵ On divergent practices between the NCAs of different NCAs, see O Brook, ‘Struggling with Article 101(3) TFEU: Diverging Approaches of the Commission, EU Courts, and Five Competition Authorities’ (2019) 56 *Common Market Law Review* 121; Or Brook and Katalin J Cseres, ‘Member State Interest in the Enforcement of EU Competition Law: A Case Study of Article 101 TFEU’ In Marton Varju (ed), *Between Compliance and Particularism: Member State Interests and European Union Law* (Springer 2018).

analyses the relationship between the Directorate-General for Competition (DG Comp or Directorate) and NCAs. Section V discusses consequent fragmentation and divergence between the practices of various NCAs. Section VI assesses the changing relationship between the Commission and NCAs in light of contemporary legislative developments. In line with its central thesis that tension between the CJEU and Commission has defined the evolution of EU competition law, the analysis of this article entails an examination of the wider institutional complex of EU competition law.

From the outset, it is worth noting that this article is primarily concerned with anticompetitive agreements prohibited by Article 101 of the Treaty on the Functioning of the European Union (TFEU). Article 101(1) sets out a blanket prohibition covering ‘all agreements between undertakings ... which may affect trade between Member States and which have as their object or effect the prevention, restriction or distortion of competition within the internal market’. Article 101(3) exempts the application of this broad prohibition where such agreements contribute ‘to improving the production or distribution of goods or to promoting technical or economic progress, while allowing consumers a fair share of the resulting benefit’. Whilst merger control and state aid also constitute key components of EU competition law, they fall outside of the purview of these provisions and are not considered in this analysis.

I. CENTRALISATION

Notwithstanding the divergence in objectives, conflict of policies, and ideological institutional tension that catalysed the decentralisation of the EU antitrust regime (with which this article is concerned), the inception of EU competition law witnessed collaboration between the CJEU and the Commission. The novel regime was ushered onto the political landscape of a nascent European Union by an ambitious Commission and an aggressive Court. Battling economic nationalism and pursuing the ‘unification imperative’ of market integration,⁶ the two institutions navigated the inauguration of an inchoate antitrust system in harmony: developing doctrine, advancing application, and concentrating power within the central authority of the

⁶David Gerber, *Law and Competition in Twentieth Century Europe: Protecting Prometheus* (Clarendon Press 1998) 347.

Commission. Whilst a ‘traditional administrative law-making model’⁷ was espoused in early EU antitrust legislation, the Court’s expansion of the application and objectives of Article 101 in the formative years of EU antitrust was instrumental in allowing the Commission to institute itself as the principal steward of EU competition law and policy. As a collaboration that shaped a formative era of EU competition law and contributed significantly to the landscape of enforcement from which proponents of decentralisation sought to depart, this institutional cooperation deserves analysis.

At the outset of EU competition law, Regulation 17/62 framed the development of competition policy as the exclusive prerogative of the Commission, conferring broad powers of investigation,⁸ prohibiting concurrent national enforcement activity⁹ and, most notably, designating exclusive competence to the Directorate in applying article 101(3) TFEU - the only mechanism by which an anti-competitive agreement may escape infringement.¹⁰ Indeed, this centralisation at the outset of EU competition law was seemingly envisaged by the Treaties¹¹ and was arguably necessitated by the breadth of the language in which the relevant provisions are couched. Owing in part to the general framework nature of the treaties and multilingual order of the EU, fundamental terms such as ‘restriction of competition’, ‘abuse’, and ‘dominant

⁷ Pablo Ibáñez Colomo, ‘Three Shifts in EU Competition Policy: Towards Standards, Decentralization, Settlements’ (2013) 20 *Maastricht Journal of European and Comparative Law* 363, 363.

⁸ Council Regulation (EEC) 17/62 First Regulation implementing Articles 85 and 86 of the Treaty [1962] OJ 13, Article 13.

⁹ *ibid* Article 9(3). Once the Commission initiated its investigation, NCAs were obliged to cease any enforcement activity on the same case.

¹⁰ *ibid* Article 9(1). Whereas Article 101(1) sets out types of prohibited agreements, Article 101(3) sets out that such agreements may escape prohibition where such concerted practice ‘contributes to improving the production or distribution of goods or to promoting technical or economic progress, while allowing consumers a fair share of the resulting benefit, and which does not: (a) impose on the undertakings concerned restrictions which are not indispensable to the attainment of these objectives; (b) afford such undertakings the possibility of eliminating competition in respect of a substantial part of the products in question’.

¹¹ Article 105(1) of the Treaty appears to envisage a system of public competition enforcement, establishing that the European Commission ‘shall ensure the application of the principles laid down in Articles 101 and 192’. The General Court has interpreted this article as constituting a specific expression of the general supervisory role of the Commission conferred by article 17(1) TEU: see Case T-24/90, *Automec v European Commission* [1992] ECR II-2223, para 74.

position' lack definition, endowing EU antitrust law with innate flexibility. Whilst such obtuse provisions would be liable to distort the uniform application of competition law in a system of decentralised application before national courts or NCAs, the concentration of enforcement competence within the Directorate allowed for 'fundamental changes' to occur 'in a seamless way'. This was through the *ex-post* development of the discipline and its doctrine in the early stages of EU competition law.¹² Indeed, in many Member States, the promotion of industrial policies generally enjoyed greater political support than the imposition of rigorous competition policy, and the concentration of this 'considerable discretionary power'¹³ of the Commission proved necessary to strengthen 'competition culture' across the Community.¹⁴

The breadth and flexibility of these provisions afforded the CJEU significant discretion in the interpretation of elementary matters of competition law. This allowed the Court to exercise 'intellectual leadership' upon which the Commission relied throughout its centralisation, consolidation, and development of authority.¹⁵ In line with its recognition of the Directorate's 'general supervisory role' in the antitrust sphere,¹⁶ the Court adopted a remarkably pervasive interpretation of the application of Article 101, bringing

¹² Pablo Ibáñez Colomo and Adriani Kalintrini, 'The Evolution of EU Antitrust Policy: 1966-2017' (2020) 83(2) *Modern Law Review* 321, 322.

¹³ European Commission, 'XXIIIrd Report on Competition Policy' (*Publications Office of the European Union*, March 25 1995) <<https://op.europa.eu/en/publication-detail/-/publication/7db4a243-39f3-4ba4-a5b7-1cb48f8ca6d3>> accessed 9 March 2023, point 190.

¹⁴ Former Director General for Competition at the European Commission Claus-Dieter Ehlermann has remarked that in the first few decades of the European Community 'the presence and interference of the Member States in their respective economies was increasing, instead of decreasing': Ehlermann (n 2) 540. In challenging this economic nationalism, EU competition law has been understood to have acted 'as an instrument that is subordinate to the achievement of [the] overarching ambition' of integrating the economies of Member States: Colomo and Kalintrini (n 12) 322.

¹⁵ Assimikis P Komninos, 'New Prospects for Private Enforcement of EC Competition Law: *Courage v Crehan* and the Community Right to Damages' (2002) 39 *Common Market Law Review* 447, 447.

¹⁶ *Automec* (n 11) para 74, in which the General Court of the CJEU held that Article 105(1) TFEU was a specific expression of the supervisory role of the Commission conferred by Article 17(1) TFEU.

almost all collusive activity within the purview of Commission enforcement.¹⁷ Whilst some agreements are exempted from enforcement under the *de minimis* doctrine,¹⁸ the expansive interpretation by the Court means that it remains ‘very rare for the EU to lack jurisdiction’.¹⁹ Article 101 may apply where all parties to an agreement are from the same Member State,²⁰ where a restrictive practice has not yet been implemented,²¹ and where an agreement has not yet had an actual impact on the market.²² Furthermore, the CJEU facilitated and supported the Commission’s expansion of the set of objectives and aims that competition law may serve, often derogating from strict economic analysis when reviewing the Commission’s infringement decisions and assessing anticompetitive effects. This was initially observed in the Court’s endorsement of the Commission’s pursuit of the ‘indigenous objective’ of market integration through competition law.²³ Whilst the Commission has continuously reiterated that this fundamental imperative and the goal of promoting competition are economically complementary,²⁴ early cases show a dearth of economic analysis in the unwavering prohibition of any practice creating barriers between Member

¹⁷ Case 56/65, *Société Technique Minière v Maschinenbau Ulm GmbH* [1966] ECR 337 249 established that Article 101 applies if it is possible ‘to foresee with a sufficient degree of probability on the basis of a set of objective factors of law or of fact that the agreement in question may have an influence, direct or indirect, actual or potential, on the pattern of trade between Member States’.

¹⁸ Agreements will not be caught if they do not have an appreciable impact on competition or on inter-state trade: see Case 5/69, *Franz Volk v J Vervaecke* [1969] ECR 295. In an official publication, the Commission has set out that agreements will not be held to appreciably restrict competition where the aggregate market share held by the parties thereto does not exceed 10% on markets where the parties are actual or potential competitors: European Commission, ‘Commission Notice on Agreements of Minor Importance Which Do Not Appreciably restrict Competition Under Article 81(1) (*de minimis*)’ [2014] OJ C291/01, para 13.

¹⁹ P Craig & G De Búrca, *EU Law: Text, Cases, and Materials* (7th edn, Oxford University Press 2020) 1059.

²⁰ Case 8/72, *Cementhandelaren BV v Commission* [1972] ECR 977.

²¹ Case T-77/92, *Parker Pen v Commission* [1994] ECR II-549.

; Case T-66/92, *Herlitz AG v Commission* [1994] ECR II-531.

²² Case 19/77, *Miller International Schallplatten GmbH v Commission* [1978] ECR 131.

²³ Ariel Ezrachi, ‘EU Competition Law Goals and the Digital Economy’ (2018) Oxford Legal Studies Research Paper No. 17/2018, 2.

²⁴ See, among others, European Commission, ‘Guidelines on the Application of Article 81(3) of the Treaty’ [2004] OJ C101/97 [13]; European Commission, ‘Guidance on the Commission’s enforcement priorities in applying Article 82 of the EC Treaty to abusive exclusionary conduct by dominant undertakings’ [2009] OJ C 45/02, paras 1.5-7.

States. Without regard to any anti-competitive effect,²⁵ all agreements prohibiting,²⁶ or even disincentivising²⁷ cross-border sales were held to be in breach of Article 101.

The economic laxity of the Court's jurisprudence subsequently allowed Article 101 to advance an even broader range of Commission objectives, with the CJEU loosening requirements of economic justification for the determination of an Article 101(3) exemption in the case of *Metro I*.²⁸ Through a teleological interpretation of Article 101, read in conjunction with Article 3 of the Treaty of Rome (EEC),²⁹ the Court held that the requirement in both articles that competition be undistorted 'implies the existence on the market of workable competition'.³⁰ According to this novel standard, the degree of competition protected under Article 101 is that which is 'necessary to ensure the observance of the basic requirements and the attainment of the objectives under the treaty'.³¹ The Court went on to hold that this requirement of 'workable competition' could be reconciled with 'the safeguarding of objectives of a different nature'.³² This judgment provided the Commission with a 'window of opportunity',³³ allowing the Directorate to grant Article 101(3) exemptions on

²⁵ See Joined Cases 56/64 and 58/64, *Établissements Consten S.à.R.L. and Grundig-Verkaufs-GmbH v Commission* [1966] ECR 299.

²⁶ Case 19/77 *Miller International Schallplatten GmbH v Commission* [1978] ECR 131.

²⁷ *GlaxoSmithKline* (n 27); see also Case 27/76, *United Brands Company and United Brands Continental BV v Commission* [1978] ECR 207.

²⁸ It is to be recalled that Article 101(3) provides for an exemption to the blanket (and widely interpreted) prohibition of collusive agreements under Article 101(1). Whereas Article 101(1) prohibits 'all agreements between undertakings, decisions by associations of undertakings and concerted practices which may affect trade between Member States and which have, as their object or effect the prevention, restriction or distortion of competition within the internal market', Article 101(3) declares this former limb inapplicable in the case of an agreement 'which contributes to improving the production or distribution of goods or to promoting technical or economic progress, while allowing consumers a fair share of the resulting benefit...'

²⁹ Article 3 EEC, whilst not included in the Lisbon Treaties, laid down a list of Community 'activities'.

³⁰ Case C-26/76, *Metro v Commission* [1977] ECR 1875, para 20.

³¹ *ibid* para 20.

³² *ibid* para 21-22.

³³ Daniel G. Goyder, *Goyder's EC Competition Law* (Oxford University Press 2009) 413.

the basis of indirect economic or non-economic objectives.³⁴ Following *Metro I*, the Commission exempted agreements on account of its objectives regarding the development of sports,³⁵ various industrial policy considerations,³⁶ the scarce allocation of resources among states³⁷ and, particularly, environmental protection.³⁸ Indeed, the Commission's 2000 guidelines³⁹ and 1995 report on competition policy⁴⁰ expressly recognise the possibility of an Article 101(3) exemption on environmental grounds. The Commission has even exceeded the parameters of the *Metro I* judgment, indistinctly applying Article 101(3) to objectives without a legal basis under the treaties, such as the survival of traditional trades.⁴¹ Indeed, this aggressive teleological reasoning and doctrinal flexibility in the Courts' economic analysis began to garner criticism;⁴² eventually leading the CJEU to demand more evidence of economic effects from

³⁴ See O Brook, 'Struggling with Article 101(3) TFEU: Diverging Approaches of the Commission, EU Courts, and Five Competition Authorities' (2019) 56 Common Market Law Review 121.

³⁵ *EBU/Eurovision System* (Case IV/32.150) Commission Decision 93/403/EEC [1993] OJ L 179/22.

³⁶ *Ford/Volkswagen* (Case IV/33.814) Commission Decision 93/49/EEC [1992] OJ L 20/28.

³⁷ *International Energy Agency* (Case IV/30.525) Commission Decision 94/153/EC [1983] OJ L 376/30.

³⁸ The Commission exempted a cartel on the basis of environmental benefit in *International Fruit Container Organisation* (Case IV/34.415) Commission Decision 97/48/EC [1997] OJ L 48/2. See also *2CECED* (Case IV.F.1/36.718) Commission Decision 2000/475/EC [2000] OJ L 187/47; see also Damien Geradin, 'EC Competition Law and Environmental Protection: Conflict or Compatibility?' (2002) 2 Yearbook of European Environmental Law 117, 124.

³⁹ See European Commission 'Notice - Guidelines on the Applicability of Article 81 of the EC Treaty to Horizontal Cooperation Agreements' [2001] OJ L 3/02.

⁴⁰ European Commission, 'XXV Report on Competition Policy' (*European Commission*, 1995) <https://ec.europa.eu/competition/publications/annual_report/1995/en.pdf> accessed 9 March 2023.

⁴¹ *Grobe's distribution system* (Case IV/30.299) Commission Decision 85/44/EC [1985] OJ L 19/17, para 25-27; see also *Ideal-Standard's distribution system* (Case IV/30.261) Commission Decision 85/45/EC [1985] OJ L 20/38.

⁴² The Court was, for instance, accused of eroding a 'juridical' conception of the law: Gerber (n 6) 351.

infringement findings of the Commission,⁴³ and even annulling decisions on the basis of a lack thereof.⁴⁴

Whilst permitted by the ‘open-textured’ and ‘programmatic’ nature of the Treaties,⁴⁵ the ‘very broad’⁴⁶ interpretation of the ‘effect on trade between Member States’ requirement under Article 101(1), and the striking margin of flexibility afforded to the Commission in its application of Article 101(3) exemptions by the doctrinal laxity of the Court, demonstrates coordination between the CJEU and Commission in the development of an expansive and dynamic centralised competition regime; a coordination reflecting coinciding integrationist objectives. As a ‘motor of integration’,⁴⁷ the CJEU sought to further the Commission’s objectives of market integration, consumer welfare, and ‘maximise’ the scope⁴⁸ of EU competition law in a manner characteristic of the ‘activist’⁴⁹ CJEU in this era of its development.⁵⁰

The divergent approaches of the two institutions began to emerge, however, as the burgeoning system of EU antitrust enforcement and the limited capacity of the Directorate necessitated systematic redesign. Whilst both the Court and Commission sought to delegate enforcement duties to national actors, different models of decentralisation, and their implications for the distribution of power between the two EU institutions, gave rise to the judicial and executive tension with which this article is concerned. The following section analyses the CJEU’s attempt to facilitate the partial decentralisation of the

⁴³ See Jean-Francois Verstrynge, ‘Current Antitrust Policy Issues in the EEC: Some Reflections on the Second Generation of Competition Policy’ in Barry Hawk (ed), *1984 Corporate Law Institute: Antitrust and Trade Policy in International Trade* (Bender 1985) 673.

⁴⁴ Case 27/76 *United Brands Company and United Brands Continentaal BV v Commission* [1978] ECR 207.

⁴⁵ Renaud Dehousse, *The European Court of Justice* (MacMillan 1998) 78.

⁴⁶ Joanna Goyder and Albertina Albors-Llorens, *Goyder’s EC Competition Law* (5th edn, Oxford University Press 2009) 101.

⁴⁷ Gerber (n 6) 351.

⁴⁸ Goyder and Albors-Llorens (n 46) 101.

⁴⁹ H Rasmussen, *On Law and Policy in the European Court of Justice – A Comparative Study* (Nijhoff 1986) 298.

⁵⁰ See, for instance, the Court’s expansive interpretation of the ‘all measures of equivalent effect’ requirement in the context of freedom of movement of goods in Case 8/74, *Procureur du Roi v Dassonville* [1974] ECR 837. See also the subsequent more restrictive reading on the same issue in Joined Cases C-267 and C-268/91, *Criminal Proceedings against Bernard Keck and Daniel Mithonard* [1993] ECR I-6097.

enforcement of EU competition law. By clarifying the competence of national courts to rule on breaches of Article 101 (notwithstanding the Commission's exclusive prerogative over Article 101(3) and recognising a Union right to damages) the CJEU sought to encourage private enforcement before the national courts of Member States, to reduce systemically unimportant complaints before the Commission, and to partially relieve the Directorate's dossier.

II. DECENTRALISATION

The Commission's centralised approach led to an 'absurd discrepancy between the Commission's theoretical jurisdiction and its capacity to generate the decisions called for by its over-broad interpretation of Article [101(1)]'.⁵¹ The accession of new Member States accompanied by a marginal increase in resources strained the Commission's investigatory and enforcement capabilities.⁵² Despite this, the Commission sought to maintain its competence over the development of antitrust policy and continuously emphasised the risk of divergence arising from a disparate application of EU competition law across Member States. Indeed, prior to entertaining decentralisation, the Commission sought to alleviate its workload through a variety of approaches, including the publication of soft-law instruments,⁵³ comfort letters and the reorganisation of the Directorate.⁵⁴ Despite such efforts, long delays continued to persist⁵⁵ and the unwieldy system attracted much criticism from non-EU corporations.⁵⁶

⁵¹ Ian Forrester and Chris Norall, 'Competition Law' (1990) 10 Yearbook of European Law 407, 409.

⁵² Between 1988 and 1998, only 13% of cases were initiated of the Commission's own volition, see European Commission, 'White Paper on modernisation of the rules implementing Articles 85 and 86 of the EC Treaty' (1999) Commission Programme No 99/027, para 44.

⁵³ These were designed to assist firms. See the European Commission's Notice on Patent Licensing Agreements [1962] OJ 139/2922; Notice on Exclusive Dealing Contracts with Commercial Agents [1962] OJ 139/2921; Notice on Cooperation Agreements [1968] OJ C75/3.

⁵⁴ See James S Venit, 'Slouching Towards Bethlehem: The Role of Reason and Notification in EEC Antitrust Law' (1987) 10 Boston College International and Comparative Law Review 17, 40.

⁵⁵ Christopher Bellamy and Graham D Child, *Common Market Law of Competition* (Sweet and Maxwell, 1987) 135.

⁵⁶ See William J Kolasky, 'Antitrust Compliance Programs: The Government Perspective', Address delivered before the Corporate Compliance Conference, July 2002.

Consistent with the foregoing exercise of its leadership, the CJEU responded. By outlining and clarifying the jurisdiction of national courts to rule on breaches of Article 101 (without encroaching upon the Commission's Article 101(3) exemption monopoly), the Court of Justice sought to stimulate antitrust actions before the courts of Member States, to reduce the traffic of disputes before the Commission, and to provide the Directorate with a mechanism through which it could remit cases of little jurisprudential significance to national courts. Circumventing the Commission's concerns regarding decentralised public enforcement, the Court sought to construct a system of decentralised private enforcement complementary to the Commission's policy prerogative.

Broadly concurrent with this initiation of decentralisation was the incorporation of subsidiarity into the EU legal order through the Maastricht Treaty.⁵⁷ Introduced to counteract potential federalist shifts,⁵⁸ the principle, now governing the 'the use of Union competences',⁵⁹ is understood as 'establishing a rule [by which] issues should be addressed at the level where they can be addressed most effectively'⁶⁰ based on an assessment of both qualitative and quantitative indicators.⁶¹ Accordingly, competences should be exercised at the EU-level only where the 'objective can be better attained at European level by reason of the scale or effects of the proposed actions'.⁶² Whilst not technically binding upon EU institutions in the sphere of competition law,⁶³ the principle has, according to former Commissioner Sir Leon Brittan, 'always underpinned' EU competition policy and encouraged the Commission to 'look afresh at the institutional balance between the Community and Member State's competition authorities to determine whether the existing system of jurisdiction allocation

⁵⁷ See Article 5(2) TEU.

⁵⁸ Paul Craig, 'Subsidiarity: a Political and Legal Analysis' (2012) 50 *Journal of Common Market Studies* 72.

⁵⁹ Article 5(1) TEU.

⁶⁰ Joel Trachtman, 'L'Etat, C'est Nous: Sovereignty, Economic Integration and Subsidiarity' (1992) 33 *Harvard International Law Journal* 459, 469.

⁶¹ Article 5(1) TEU.

⁶² Carlo Petrucci, 'Subsidiarity in Directive 2014/104 EU on Damages Actions for Breach of EU Competition Law' (2017) 23(2) *European Public Law* 395, 400.

⁶³ Competition law is an area which falls within the exclusive competence of the EU and thus Article 5(2) TEU does not apply.

can be further refined'.⁶⁴ However, as the DG Comp's docket expanded, this theoretical ambition became a pragmatic necessity.⁶⁵ The notion of subsidiarity provides a convenient prism through which the decentralisation of EU competition law and the involvement of national courts in the enforcement of EU antitrust law can be assessed.

In accordance with the definition of subsidiarity under the Treaties, such national enforcement must be 'efficient, effective and adequate' in order for national courts to 'sufficiently achieve the objectives' of Articles 101 and 102.⁶⁶ The Commission's centralised regime prevented such enforcement due to: (i) the possibility of contradictory rulings between the national courts and the Commission, and (ii) the absence of effective relief available before national courts. By redefining the 'division of competence'⁶⁷ between national courts and the Commission, the CJEU addressed both obstacles, thereby empowering national courts, relieving the Commission of inconsequential cases, and promoting subsidiarity in the context of EU antitrust law.

(i) Contradictory Rulings

Whilst the Commission's monopoly over the granting of Article 101(3) exemptions constituted the biggest obstacle to national enforcement of Article 101, various mechanisms under Regulation 17 preserving the Commission's capacity to intervene in the course of national proceedings further deterred suits before the courts of Member States. National courts lost the jurisdiction to adjudicate claims as soon as the Commission initiated its procedures of negative clearance,⁶⁸ termination of infringement,⁶⁹ or exemption.⁷⁰ Considering the

⁶⁴ Speech by Sir Leon Brittan in Roger P Alford, 'Subsidiarity and Competition: Decentralized Enforcement of EU Competition Laws' (1004) 27 *Cornell International Law Journal* 2, 272.

⁶⁵ *ibid.*

⁶⁶ *ibid* 292.

⁶⁷ Case C-234/89 *Delimitis v. Henninger Bräu* [1991] ECR I-935, para 47.

⁶⁸ Regulation 17 (n 8) Article 2.

⁶⁹ *ibid* Article 3.

⁷⁰ *ibid* Article 6. Exemptions under Articles 2, 3 and 6 were cumbersome, and took considerable time. The Commission rarely issued more than a dozen each year: See Frances Graupner, 'Commission Decision-Making on Competitive Questions' (1973) 10 *Common Market Law Review* 291 and Valentine Korah, 'The Rise and Fall of Provisional Validity - The Need for a Rule of Reason in EEC Antitrust' (1981) 3

possibility of exemption by the Commission following a national court decision voiding an agreement under Article 101(1) or (2)⁷¹ and the ‘material injustice’ arising therefrom,⁷² it had previously been held that notification to the Commission had a suspensive effect on national proceedings, greatly disincentivising claims before national courts.⁷³ This ‘suspensive effect’ was, however, subsequently lifted by the Court of Justice.⁷⁴ The Court thereafter continued to deconstruct barriers to national private enforcement by (i) clarifying the ‘division of competence’⁷⁵ between the Commission’s monopoly in Article 101(3) and national courts’ jurisdiction, and (ii) bypassing recourse to Article 101(3) by undermining a strict approach in the application of Article 101(1).

In *Delimitis*, the CJEU distinguished between clearly prohibited agreements with regard to which national courts could make findings, and agreements which may merit an exemption under Article 101(3). In response to this, national courts were obliged to ‘stay proceedings or adopt interim measures’ until the Commission reached its verdict.⁷⁶ This reasoning differentiated antitrust cases before national courts along a continuum of complexity, ranging from the ‘pedestrian’⁷⁷ to the ‘highly complex’.⁷⁸ Similarly, in *Automec*, the Court, whilst emphasising the role of the Commission as an administrative authority acting in the ‘public interest’,⁷⁹ upheld the Commission’s refusal to investigate certain complaints where an adequate remedy was available

Northwestern Journal of International Law and Business. However, the prospect of such a decision nonetheless disincentivised litigation: see Richard Whish and Brenda Sufrin, ‘Article 85 and the Rule of Reason’ (1987) 7(1) Yearbook of European Law 1.

⁷¹ Article 101(2) TFEU provides that any agreement prohibited by article 101 shall be deemed automatically void. This has been termed the ‘Euro-defence’ in competition law: see Richard Whish, *Competition Law* (4th edn, Oxford University Press 2001) 266.

⁷² European Commission, ‘The Application of Articles 85 and 86 of the EC Treaty by National Courts in the Member States’ (*European Commission*, 18 July 1997) <<https://op.europa.eu/en/publication-detail/-/publication/03d808bf-3928-4793-bcfc-72bb63da27dd>> accessed 9 March 2023, 281.

⁷³ John Temple Lang, ‘EEC Competition Actions in Member States’ Courts - Claims for Damages, Declarations and Injunctions for Breach of Community Antitrust Law’ (1984) 7 Fordham International Law Journal 389, 407.

⁷⁴ Case 48/72, *SA Brasserie de Haecht v. Wilkin-Janssen* (No. 2) [1973] ECR 77.

⁷⁵ *Delimitis* (n 67) para 47.

⁷⁶ *ibid* para 52.

⁷⁷ Alford (n 64) 281.

⁷⁸ Temple Lang (n 73) 246.

⁷⁹ *Automec* (n 11) para 63.

before a national court. Furthermore, the Court articulated a test determining the circumstances in which the Commission should launch an investigation, including a requirement for the directorate to have regard to whether there is ‘sufficient Community interest’.⁸⁰ While this approach was welcomed by the Commission as providing authority for its remittal of burdensome and jurisprudentially insignificant complaints to national courts,⁸¹ it also defined, and arguably restricted, the Commission’s role in a hierarchy of antitrust adjudication to assessing ‘important questions of legal precedent’ and necessitated the application of the subsidiarity principle in the Directorate’s practice.⁸²

Whilst this development in the case law was of indisputable benefit to the Commission by increasing antitrust enforcement before national courts⁸³ and decreasing the Directorate’s workload,⁸⁴ it is worth noting the Court’s concentration on the importance of facilitating parties’ claims, and implicit emphasis on ensuring adequate compensation. In *Automec*, the Court relied upon the postulation of the Treaty ‘that national law gives the court power to preserve the rights of enterprises which are victims of anti-competitive practices’.⁸⁵ In *Brasserie de Haecht*, the Court’s abrogation of the suspensive effect of notification to the Commission was based on national courts’ ‘obligation of deciding on the claims of interested parties’.⁸⁶ The CJEU’s reasoning in *Delimitis* and *Automec* appears to be firmly undergirded by an endorsement of subsidiarity and an emphasis on the rights of antitrust victims; motives difficult to reconcile with the Commission’s monopoly over Article 101(3). This reasoning went so far as to prompt the Commission to publish guidance to national courts in the months following *Automec* clarifying the deference owed by national courts to

⁸⁰ *ibid* para 98.

⁸¹ Speech by Sir Leon Brittan in Alford (n 64).

⁸² Alford (n 64).

⁸³ European Commission, ‘Thirteenth Report on Competition Policy’ (*European Commission*, 1 August 1996).

<<https://op.europa.eu/en/publication-detail/-/publication/161bd425-29e6-4ac0-9b3d-0766ccdda8cd>> accessed 9 March 2023, 217.

⁸⁴ European Commission, ‘Sixteenth Report on Competition Policy’ (*European Commission*, 18 April 1990).

<<https://op.europa.eu/en/publication-detail/-/publication/21ec96eb-ff7d-49e1-bfd3-6c56a4846e5c>> accessed 9 March 2023, 52.

⁸⁵ *Automec* (n 11) para 93.

⁸⁶ *Brasserie da Haecht* (n 74) para 11.

the former decisions,⁸⁷ comfort letters,⁸⁸ and exemption contemplation⁸⁹ of the Commission.

Furthermore, whilst the foregoing analysis concerns exemptions under the third limb of Article 101, it is also of note that agreements may also escape Article 101 by failing to come within the provisions of Article 101(1) to begin with.⁹⁰ Whilst the broad terms of the treaty and generous interpretation afforded thereto⁹¹ encompassed the vast majority of collusive arrangements, it seems that the court covertly undermined this broad, indiscriminate application of Article 101(1) by gravitating towards a ‘rule of reason’ approach in the application of Article 101.⁹² This approach imports an assessment of the pro- and anti-competitive effects of an agreement into the initial determination of the satisfaction of the test of Article 101(1) rather than allowing such effects to weigh in favour of an Article 101(3) exemption at the subsequent stage of analysis, undermining the Commission’s exclusive competence over an analysis crucial to its policy prerogative and threatening to afford unprecedented discretion to national courts.⁹³ Although the merits of such an approach have garnered both academic scepticism and support,⁹⁴ the consistency of this doctrinal laxity with the Treaties is dubious.

The satisfaction of Article 101(1) requires that agreements have ‘as their object or effect the prevention, restriction or distortion of competition within the internal market’. Deeming these criteria to be disjunctive,⁹⁵ earlier decisions of the CJEU strictly delineated the boundary between anti-competitive

⁸⁷ European Commission, ‘Notice on Cooperation Between National Courts and the Commission in Applying Articles 85 and 86 of the EEC Treaty’ [1993] OJ C 39/6.

⁸⁸ *ibid.*

⁸⁹ *ibid.*

⁹⁰ It is to be recalled that Article 101(1) contains the general prohibition of anticompetitive agreements, whilst Article 101(3) provides for the exemption of such agreements on various (and broadly couched) grounds.

⁹¹ See *Société Technique Minière* (n 17), the test of which is set out above.

⁹² See *Standard Oil v US* 337 U.S. 293 (1949) (US Supreme Court).

⁹³ At the emergence of this approach, commentators observed the potential for a ‘rule of reason’ side-step to Article 101 to stimulate private claims before national courts: see Korah (n 70) 350.

⁹⁴ See E Fox, ‘The Modernization of Antitrust: A New Equilibrium’ (1981) 66 Cornell Law Review 1140; Korah (n 70) 310; Whish and Sufrin (n 70).

⁹⁵ *Société Technique Minière* (n 17).

agreements ‘by object’ and ‘by effect’.⁹⁶ The former type of agreement is prohibited ‘by its very nature’,⁹⁷ to be determined by reference to its content, objectives, and economic and legal context.⁹⁸ In the case of the latter type of agreement, where anti-competitive quality is not evident from its object,⁹⁹ the Court has incorporated economic analysis of the effects of the agreement on the market when determining the satisfaction of Article 101(1),¹⁰⁰ utilising Article 101(1) as a provision of economic assessment rather than a provision of jurisdiction.¹⁰¹ Whilst the ‘rule of reason’ was deemed ‘difficult to reconcile with the rules prescribed by [Article 101(1)]’ by the Court of First Instance,¹⁰² this ‘more nuanced analysis’¹⁰³ under Article 101(1) became prominent in the jurisprudence of the European courts, with agreements imposing selective distribution schemes,¹⁰⁴ exclusive licences,¹⁰⁵ non-compete clauses,¹⁰⁶ exclusive purchasing obligations,¹⁰⁷ and even territorial protection in franchising arrangements,¹⁰⁸ evading Article 101(1).¹⁰⁹ However, despite the emergence and academic support of this more flexible approach,¹¹⁰ the Commission failed to follow suit, and continued ‘mechanically applying’¹¹¹ Article 101 to agreements

⁹⁶ *ibid*; *Consten Grundig* (n 25).

⁹⁷ Case C-67/13 P, *Groupement des Cartes Bancaires v European Commission* [2014] EU:C:2014:2204.

⁹⁸ *GlaxoSmithKline* (n 27).

⁹⁹ *Brasserie de Haecht* (n 74); *Volk* (n 18).

¹⁰⁰ *Consten Grundig* (n 25).

¹⁰¹ *Goyder* (n 46) 111.

¹⁰² Case T-112/99 *Metropole Télévision v Commission* [2001] ECR I-7945; cf Case T-374/94, *European Night Services v Commission* [1998] ECR II-03141.

¹⁰³ James S Venit, ‘Brave New World: The Modernization and Decentralization of Enforcement under Articles 81 and 82 of the EC Treaty’ (2003) 40(3) Common Market Law Review 545, 551.

¹⁰⁴ Case 75/84 *Metro SB-Großmärkte GmbH & Co. KG v Commission* [1986] ECR 3021.

¹⁰⁵ Case 258/78 *Nungesser KG and Kurt Eisele v Commission* [1982] ECR 2015.

¹⁰⁶ Case 42/84 *Remia BV and others v. Commission* [1985] ECR 2545.

¹⁰⁷ Case C-250/92 *Gottrup-Klim e.a. Grovareforeninger v Dansk Landbrugs Grovarelskab AmbA* [1994] ECR 5641.

¹⁰⁸ Case 161/84 *Pronuptia de Paris GmbH v Pronuptia de Paris Irmgard Schillgallis* [1986] ECR 353.

¹⁰⁹ Note, however, that the Court maintained a strict application of Article 101 in certain cases, such as in regard to agreements seeking to resurrect territorial borders between Member States: see Case 258/78 *LC Nungesser KG and Kurt Eisele v Commission* [1982] ECR 2015.

¹¹⁰ See Valentine Korah, ‘Critical Comments on the Commission’s Recent Decisions’ (1987) 12(1) European Law Review 18.

¹¹¹ Whish and Sufrin (n 70) 22.

which did not, in reality, distort competition.¹¹² Indeed, this gravitation towards a ‘rule of reason’ undermined the Commission’s monopoly, and, as will be discussed below, was expressly recognised as a threat to the utility of Article 101(3) in the Commission’s White Paper on Modernisation — a document representing the Directorate’s concession to the pressure of decentralisation.¹¹³ Akin to the Commission’s refusal to pursue a ‘rule of reason’ style of analysis in the context of Article 101(1), the Directorate similarly disliked the disruption posed by the recognition of a Union right to damages to the successful Leniency scheme. As discussed in Section IV, the cumulative effect of these developments instigated systemic redesign from the Commission.

(ii) Effective Relief

The foregoing mechanism for remittal to national courts is, however, based upon the availability of effective relief. Whilst required in principle by the uniform application of Community law and subsidiarity, the Commission’s ability to remit claims to be adjudicated by national courts established in *Automec* is postulated upon effective judicial enforcement in Member States. As stated by Advocate General Edwards in that case, the Commission cannot simply ‘repeat a ritual formula to the effect that relief is available in the national courts’.¹¹⁴ In the seminal 2001 *Courage v Crehan* case, the ECJ established the right of ‘any individual’ to bring an action to obtain compensation for harm suffered as a result of anti-competitive practice.¹¹⁵ Considered ‘strictly linked’¹¹⁶ to the effective application of EU law, the ECJ required national courts to ensure the protection of this right. This promotion of the private enforcement of

¹¹² See, in the context of Research and Development agreements, *Vacuum Interrupters* agreement [1977] 1 CMLR D67 and *Rockwell/Iveco* [1983] CMLR 709. See, however, the less indiscriminate approach of the Commission in the context of franchising agreements: *Pronuptia* (Case IV/30937) Commission Decision 87/17/EEC [1986] OJ L 013.

¹¹³ Commission Programme No 99/027 (n 52). It should be noted that, recently, the Court has undermined the distinction between object and effect.

¹¹⁴ Opinion of Advocate General Edwards in *Automec* (n 11) para 116.

¹¹⁵ Case C-453/99 *Courage v Crehan* [2001] ECR I-6297. This followed the prior confirmation of the CJEU that Articles 101 and 102 create relations between and rights for individuals in Case 127/73 *Belgische Radio en Televisie and Society Belge des Auteurs, Compositeurs et Editeurs de Musique v SV SABAM and NV Fonior* [1974] ECR 313.

¹¹⁶ Roberto Cusotta, ‘Some Considerations on the Last Development on Antitrust Damages Actions and Collective Redress in the European Union’ (2014) 10(1) Competition Law Review 81, 85.

competition law proved incongruous with mechanisms of public enforcement, exposing divergent priorities and objectives of the Court and Commission in their navigation of decentralisation.

Whilst the Commission, through the publication of various reports,¹¹⁷ studies,¹¹⁸ and notices,¹¹⁹ had timidly emphasised the desirability of private damages actions in the sphere of antitrust for decades prior to *Courage*, the interaction of this right with its ‘Leniency’ mechanism proved difficult. Introduced in 1996¹²⁰ and subsequently adapted,¹²¹ the Leniency Notice allows cartel participants to provide competition authorities with ‘presentations regarding that participant’s knowledge of, and role in, the cartel’ in exchange for immunity or a fine reduction.¹²² This programme has been particularly effective in the prosecution of cartels,¹²³ yielding staggering fine revenues while requiring less investigative effort from the Directorate. The appeal of the leniency scheme was undermined, however, by the emergence and subsequent clarification of a Union right to damages. Whilst an undertaking may escape fines from the Commission in lieu of cooperation, the leniency programme could not exempt private suits. A thread of jurisprudence following *Courage* exacerbated this effect.

In *Pfleiderer*, the Court of Justice held that EU law does not preclude national law from granting a litigant access to documents compiled under the leniency programme.¹²⁴ The Court further held that the act of balancing the

¹¹⁷ Rapport Deringer, Assemblée Parlementaire Européenne, Documents de Seance 1961–1962, 7 Septembre 1961, Document 57, para 123.

¹¹⁸ See the study examining the available remedies in the national laws of Member States for competition law infringement: *La Réparation des Conséquences Dommageables d’une Violation des Articles 85 et 86 du Traité Instituant la CEE*, Serie Concurrence No. 1 (Brussels, 1966).

¹¹⁹ European Commission Notice C39/6 (n 87).

¹²⁰ First adopted in the European Commission, ‘Commission Notice on the non-imposition or reduction of fines in cartel cases’ [1996] OJ C 207/4.

¹²¹ Replaced in 2002 ([2002] OJ C 45/3). The current policy is in the new notice published in 2006 (Notice on Immunity from fines and reduction of fines in cartel cases [2006] OJ C 298/17).

¹²² Article 2(15) of Directive 2014/104/EU of the European Parliament and of the Council of 26 November 2014 on certain rules governing actions for damages under national law for infringements of the competition law provisions of the Member States and of the European Union [2014] OJ L349/1 (Damages Directive).

¹²³ See Declan Walsh, ‘Carrots and Sticks - Leniency and Fines in EC Cartel Cases’ [2008] 1 ECLR 30.

¹²⁴ Case C-360/09 *Pfleiderer AG v BundesKartellamt* [2011] ECR I-5161.

right to compensation against the imperative of protecting the leniency programme's effectiveness fell within the competence of national courts. Moreover, in *Donau Chemie*, the Court held that EU law precludes national law from requiring all parties to consent to the disclosure of documents lodged by a public authority in the context of a national leniency programme 'without leaving any possibility for the national courts of weighing up the interests involved'.¹²⁵ The Court based this reasoning on the obligation of undertakings to 'compensate for the harm resulting from an infringement' of Article 101.

Following *Pfleiderer*, national courts granted access to leniency files contrary to the *amicus curiae* advice of the Commission,¹²⁶ and the General Court permitted access to documents compiled in leniency cases.¹²⁷ This subversion of the confidentiality and protection of leniency material has impeded the effectiveness of the Commission's leniency programme,¹²⁸ and remains an issue for practitioners to this day. Some studies estimate a subsequent 50% reduction in leniency applications even following the Damages Directive (discussed below)¹²⁹ and 83% of practitioners have reported a decreased interest amongst clients in applying for leniency.¹³⁰

It may thus be seen that by permitting the delegation of trivial cases to national courts, circumscribing the scope of the Commission's exemption monopoly, undermining the exclusive function of this exemption prerogative, and promoting litigation amongst victims of antitrust infringement, the first movement towards decentralisation of the enforcement of EU competition law occurred through successive doctrinal advancements of the CJEU. This was

¹²⁵ Case C-536/11 *Bundeswettbewerbsbehörde v Donau Chemie AG and Others* [2013] ECLI:EU:C:2013:366.

¹²⁶ See *National Grid Electricity Transmission Plc v ABB Ltd and Others* [2012] EWHC 869 (Ch).

¹²⁷ See Case T-437/08 *CDC Hydrogen Peroxide v European Commission* [2011] ECR II-8251. See also Case C-365/12 P *Energie Baden-württemberg AG v European Commission* [2014] ECLI:EU:C:2014:112.

¹²⁸ Johan Ysewyn and Siobhán Kahmann, 'The Decline and Fall of the Leniency Programme in Europe' (2018) 1 *Concurrences* 44.

¹²⁹ Global Competition Review, 'European Union's Directorate-General for Competition: Rating Enforcement 2016/2017' (September 2017).

¹³⁰ Wouter Wils, 'Private Enforcement of EU Antitrust Law and Its Relationship with Public Enforcement: Past, Present and Future' (2017) 40(1) *World Competition: Law and Economics Review* 3, 42.

underpinned by notions of compensatory rights and an implicit endorsement of subsidiarity, but informed by the contemporaneous exigency of the Commission's unsustainable caseload. In accordance with the widely accepted view that extra-legal considerations have considerably influenced the pace of the CJEU's integrationist jurisprudence,¹³¹ the Court exercised its role as the 'principal engine in the integration process'¹³² in engaging national actors in the enforcement of EU competition law despite persistent aversion within the Commission to decentralisation across NCAs.¹³³ Similarly seen in the context of its promotion of a right to equal pay under Article 157 TFEU in response to national legislative lacunae,¹³⁴ and recognition of a right to establishment under Article 49 TFEU in the context of inert Union harmonisation efforts,¹³⁵ the CJEU attempted to remodel EU competition law through the recognition of a directly enforceable right under Article 101 and design of an effective remittal mechanism. Whilst ostensibly in line with the Commission's aspirations for the increased engagement of national courts in the enforcement of competition law,¹³⁶ the pernicious effect of private enforcement upon the Commission's

¹³¹ Dehousse (n 45) 78. Whilst some integration theorists view the case law of the court as merely giving effect to broader political objectives (Geoffrey Garrett, 'International Cooperation and Institutional Choice: The European Community's Internal Market', (1992) 41 *International Organization* 533; Andrew Moravcsik, 'Preferences and Power in the European Community: A Liberal Intergovernmentalist Approach' (1993) 31(4) *Journal of Common Market Studies* 473) such assertions seem reductive. Whilst recognising that the Court is, indeed, an autonomous judicial institution (see Pierre Pescatore, 'The Doctrine of "Direct Effect": An Infant Disease of Community Law' (1983) 40 (2) *European Law Review* 135) the influence of 'extra-legal considerations' upon the European judiciary is widely accepted (albeit to various extents) (see Renaud Dehousse and Joseph H.H. Weiler, 'The Legal Distinction' in W Wallace (ed) *The Dynamics of European Integration* (Pinter 1990) 242).

¹³² Dehousse (n 45) 76.

¹³³ See Alexander Schaub, 'EC Competition System: Proposals for Reform' (1998) 22 *Fordham International Law Journal* 853, 876.

¹³⁴ In Case 43/75 *Defrenne v Sabena* [1976] ECR 455, a case concerning a lacuna in Belgian rules, the Court held that Article 157 TFEU, requiring equal pay for men and women, does not only impose obligations on Member States, but also confers a right on individuals which can be invoked before national courts.

¹³⁵ In Case 2/74 *Reyners v Belgian State* [1974] ECR 631, the Court adopted a similar tactic to that in *Defrenne* in the context of freedom of establishment. In response to a failure of the EU legislators to harmonise measures concerning non-salaried activities, the CJEU held that Article 49 TFEU contains a prohibition on discrimination on the basis of nationality which could be invoked by individuals before national courts.

¹³⁶ See Commission Notice on cooperation between national courts and the Commission in applying Articles 85 and 86 of the EEC Treaty, OJ [1993] C 39/6.

‘especially important’¹³⁷ leniency mechanism (contrary to the wishes of the Directorate and its central objectives of efficiency and consumer welfare¹³⁸) and the threat posed to the Commission’s policy monopoly over Article 101(3) necessitated systematic overhaul. In a manner reminiscent of the Court’s catalysis and acceleration of the 1972 merger regulation,¹³⁹ the foregoing jurisprudence culminated in the novel regime of Modernisation.

III. MODERNISATION

The promotion of private enforcement, the impediment to public implementation, and the reallocation of authority posed by the Court’s recognition of a Union right to damages necessitated systematic overhaul from the Commission. With ‘parameters for the future choices of the legislature’ irreversibly oriented by such jurisprudence, the scope for reform was limited.¹⁴⁰ Though previously deemed ‘taboo’,¹⁴¹ the Commission chose to embrace the ‘heresy’¹⁴² of decentralisation on its own terms, adroitly fashioning a novel regime in which it remained a ‘dominant voice and the control organ’.¹⁴³

¹³⁷ Neelie Kroes, ‘Tackling Cartels – a never-ending task. Anti-Cartel Enforcement: Criminal and Administrative Policy’ (2009) (SPEECH/09/454).

¹³⁸ Wils (n 130) 42.

¹³⁹ In response to Member State objections to an EU merger regulation regime and a lack of Treaty provision for mergers and acquisition control, the Court, taking a teleological interpretation of Article 102 TFEU in Case 6/72, *Continental Can v Commission* [1975] ECR 495 permitted the Commission’s use of Article 102 to address market dominance arising from mergers. Subsequently, the Court held in Joined Cases 142 and 156/84 *Phillip Morris* [1987] ECLI:EU:C:1987:490 that mergers arising from agreements between companies might be prohibited by Article 101 TFEU. The uncertainties arising from this jurisprudence afforded ‘sufficient salience’ to the necessity of an EU-wide merger regime and precipitated the 1972 EU Merger Regulation: Simon Bulmer, ‘Institutions and Policy Change in the European Communities: The Case of Merger Control’ (1994) 72(3) Public Administration 423, 432.

¹⁴⁰ Roberto Cusotta, ‘Some Considerations on the Last Development on Antitrust Damages Actions and Collective Redress in the European Union’ [2014] 10(1) Comparative Law Review 81, 99.

¹⁴¹ Alexander Schaub, ‘EC Competition System: Proposals for Reform’ (1998) 22 Fordham International Law Journal 853, 876.

¹⁴² Ehlermann (n 2) 538.

¹⁴³ David Gerber, ‘Two Forms of Modernization in European Competition Law’ (2007) 31 Fordham International Law Journal 1235, 1242.

In 1999, the Commission proposed a complete devolution of the Article 101(3) exemption in its Modernisation White Paper, empowering national authorities to apply Article 101 TFEU in its entirety.¹⁴⁴ Ultimately leading to the enactment of Regulation 1/2003, this document, promoting the diffusion of power across the vast network of national competition authorities, has been described as ‘the most important policy paper the Commission has ever published in more than 40 years of [EU] competition policy’.¹⁴⁵ However, whilst ending the procedural bifurcation of Article 101, this ‘big bang’¹⁴⁶ may not have been the ‘subsidiarity revolution’ envisaged by certain commentators.¹⁴⁷ Indeed, in line with the Commission’s ‘almost religious belief’ in the centralised concentration of power,¹⁴⁸ this ‘brave new world’ of competition enforcement¹⁴⁹ has been subsequently interpreted as amounting to an ‘extraordinary coup’,¹⁵⁰ allowing the Commission to consolidate and reaffirm its power. This section seeks to establish that the Commission achieved such a reassertion of authority despite decentralised enforcement in three ways: The Directorate (i) prioritised public enforcement to the detriment of private claims, (ii) limited the assessment of ‘benefit’ under Article 101(3) to objective economic analysis, and (iii) developed the European Competition Network.

(i) Prioritising Public Enforcement

Through the enactment of Regulation 1/2003 and the supplementary 2014 Damages Directive, the CJEU’s promotion of national courts as agents for competition enforcement has been counteracted by legislative delegation of Article 101(3), a devolution positioning national competition authorities as the ‘principal mechanisms for applying and enforcing community law’.¹⁵¹ Although

¹⁴⁴ Commission Programme No 99/027 (n 52).

¹⁴⁵ Stephen Wilks, ‘Agency Escape: Decentralization or Dominance of the European Commission in the Modernization of Competition Policy?’ (2005) 18(3) *Governance* 431, 435.

¹⁴⁶ Ivo Van Bael, ‘A “Big Bang” in EC Antitrust Enforcement Procedure’, in P. Demaret, I. Govaere and D. Hanf (eds), *30 Years of European Legal Studies at the College of Europe* (PIE Peter Lang 2005).

¹⁴⁷ Alford (n 64) 296.

¹⁴⁸ Ehlermann (n 2).

¹⁴⁹ Venit (n 103) 545.

¹⁵⁰ Wilks (n 145) 437.

¹⁵¹ David Gerber, ‘Modernising European Competition Law: A Development Perspective’ (2001) 22 *European Competition Law Review* 122, 128.

this novel regime also appears to empower national courts, provisions protecting leniency material and prioritising public enforcement have sought to limit national actions to follow-on private suits, rendering damages for antitrust violation a mere supplement to fines imposed by NCAs and the Commission.

At first glance, the Damages Directive ostensibly facilitates private actions. The findings of NCAs are accorded an enhanced probative value before courts. Article 9 renders an infringement finding of an NCA irrefutable proof of such infringement before a national court,¹⁵² and ‘*prima facie*’ proof before the courts of other Member States.¹⁵³ Furthermore, Article 17 imposes a rebuttable presumption that collusive activity causes harm.¹⁵⁴ Indeed, the Directive was drafted in consideration of the lack of effective rules on private enforcement in some Member States,¹⁵⁵ the general paucity of judgments resulting in awards,¹⁵⁶ and the inconsistency of such awards across Member States.¹⁵⁷ Furthermore, empirical evidence suggests that the Damages Directive has succeeded in promoting private actions; with 239 such cases coming before national courts in 2019 compared to a mere 50 in 2014.¹⁵⁸

However, practice also suggests that the vast majority of these actions merely follow the infringement findings of public authorities.¹⁵⁹ Whilst the Directive facilitates ‘follow-on’ litigation, its protection of leniency material has the opposite effect on ‘stand-alone’ suits. National courts are forbidden from

¹⁵² Damages Directive Article 9(1).

¹⁵³ *ibid* Article 9(2).

¹⁵⁴ *ibid* Article 17.

¹⁵⁵ Commission, ‘Commission Staff Working Document Impact Assessment Report, Damages actions for breach of the EU antitrust rules accompanying the proposal for a Directive of the European Parliament and of the Council on certain rules governing actions for damages under national law for infringements of the competition law provisions of the Member States and of the European Union’ COM (2013) 404 final, 23.

¹⁵⁶ Denis Waelbroeck, Donald Slater, and Gil Even-Shoshan, ‘Study on the Conditions of Claims for Damages in Case of Infringement of EC Competition Rules: Comparative Report’ [2004] Ashurst Report.

¹⁵⁷ Barry Rodger, *Competition Law: Comparative Private Enforcement and Collective Redress Across the EU* (Wolters Kluwer 2014).

¹⁵⁸ Jean-François Laborde, ‘Cartel damages actions in Europe: How courts have assessed cartel overcharges’ (2019) 4 *Concurrences* 7; See also Lukas Rengier, ‘Cartel Damages Actions in German Courts: What the Statistics Tell Us’ (2020) 11(1-2) *Journal of European Competition Law & Practice* 72.

¹⁵⁹ Rodger (n 157).

ordering the disclosure of leniency statements and settlement submissions,¹⁶⁰ and all other information prepared by parties during the course of a national competition authority's proceedings cannot be released until after the conclusion of the investigation.¹⁶¹ This is consistent with the Commission's concentration on the deterrent potential of private actions rather than their compensatory benefit.¹⁶² Indeed, before altering its stance in the White Paper and subsequent press releases,¹⁶³ the Commission's Green paper recognised that both public and private enforcement serve to 'deter anti-competitive practices forbidden by antitrust law'¹⁶⁴ and the Director General of the Directorate stated that 'compensation of victims should not be seen as an end in itself, but part of an overall strategy to enhance deterrence'.¹⁶⁵ Additionally, former Commissioner for Competition Neelie Kroes continuously referred to private litigation as a 'complementary'¹⁶⁶ system of enforcement which ought to work 'in parallel' with its public counterpart.¹⁶⁷ Whilst the compensation of victims of antitrust violations formed the basis of the CJEU's decentralisation efforts, it has been reframed as merely a beneficial by-product of the Commission's redefined system of public enforcement.

In addition to protecting the 'inherently superior' system of public enforcement in EU competition law,¹⁶⁸ this limitation of antitrust damages to follow-on awards, and the concomitant relegation of the role of private suits, has diminished the influence of both national courts and, by a consequential

¹⁶⁰ Damages Directive Article 6(6). Note also the broad definition of leniency statements in article 2(16).

¹⁶¹ *ibid* Article 6(5) and 7(2).

¹⁶² See Firat Cengiz, 'Antitrust Damages Actions: Lessons From American Indirect Purchasers' Litigation' (2010) 59 *International and Comparative Law Quarterly* 39.

¹⁶³ European Commission, 'Commission Recommends Member States to Have Collective Redress Mechanisms in Place to Ensure Effective Access to Justice' (Press release IP/13/524) 2.

¹⁶⁴ Commission, 'Green Paper - Damages actions for breach of the EC antitrust rules' COM (2005) 672 final 3.

¹⁶⁵ OECD, *Private Remedies*, DAF/COMP(2006) 34, 271.

¹⁶⁶ Neelie Kroes, 'Enhancing Actions for Damages for Breach of Competition Rules in Europe' [2005] Speech/05/533 3.

¹⁶⁷ Mario Monti, 'Private litigation as a key complement to public enforcement of competition rules and the first conclusions on the implementation of the new Merger Regulation' [2004] Speech/04/403.

¹⁶⁸ Wouter PJ Wils, 'Should Private Antitrust Enforcement be Encouraged in Europe?' (2003) 26(3) *World Competition* 473, 484.

restriction of the scope of Article 267 references, the CJEU.¹⁶⁹ Not only does the protection of leniency materials greatly reduce the chance of stand-alone damages being awarded,¹⁷⁰ but the irrefutable probative value of a domestic NCAs' findings before a national court¹⁷¹ limits the role of the national judges to the mere assessment and calculation of damages, diminishing the role of jurisprudence in the development of competing policy and strategy. Indeed, such rules have been interpreted as operating to thwart artful judicial reasoning by 'limit[ing] national policymakers' room to manoeuvre to introduce more effective rules that could facilitate damages claims'.¹⁷² Furthermore, this residual jurisdiction of national courts has been encroached upon by soft-law and *amicus curiae* guidance by public enforcement authorities. The Commission has issued guidance on the calculation of quantum,¹⁷³ passing-on,¹⁷⁴ and confidentiality;¹⁷⁵ and has offered *amicus curiae* advice in cases concerning limitation periods,¹⁷⁶ the parallel application of EU and national law,¹⁷⁷ and optimal fining.¹⁷⁸

¹⁶⁹ Article 267 TFEU provides a mechanism by which national courts of Member States may refer questions of EU law to the CJEU. Note that the ECJ may also review the infringement decisions of public enforcement actors under Article 263.

¹⁷⁰ Wils (n 168).

¹⁷¹ Damages Directive Article 9(2). See also recital 34 which states that this binding effect 'cover[s] only the nature of the infringement and its material, personal, temporal and territorial scope as determined by the competition authority or review court'. Note also that the original proposal for the Damages Directive rendered the decisions of foreign NCAs binding before national courts.

¹⁷² Carlo Petrucci, 'Subsidiarity in Directive 2014/104 EU on Damages Actions for Breach of EU Competition Law' (2017) 23(2) European Public Law 395, 408.

¹⁷³ European Commission, Communication from the Commission on quantifying harm in actions for damages based on breaches of Arts 101 or 102 of the Treaty on the Functioning of the European Union [2013] OJ C167/19.

¹⁷⁴ European Commission, Communication from the Commission - Guidelines for national courts on how to estimate the share of overcharge which was passed on to the indirect purchaser [2019] OJ C267/4, 43.

¹⁷⁵ European Commission, Communication from the Commission – Communication on the protection of confidential information by national courts in proceedings for the private enforcement of EU competition law [2020] OJ C242/1.

¹⁷⁶ *Morgan Advanced Materials v. Deutsche Bahn* [2014] UKSC 24.

¹⁷⁷ *E.-P. v. Slovak Competition Authority*, 8Sžhp/1/2012 [2012] (Slovak Supreme Court).

¹⁷⁸ *X BV v. Inspecteur Belastingdienst*, case 06/00252 [2010] (Gerechtshof Amsterdam, the Netherlands).

Many laud the Damages Directive as ‘exploit[ing] the synergies’¹⁷⁹ and ensuring optimal interaction between public and private enforcement. However, an analysis of its provisions in the context of the institutional tension underlying the decentralisation of EU antitrust law also reveals that the Commission and EU legislators have harnessed the efforts of private litigants to their own ends, restricting damages actions to a complementary sanctioning function. Private enforcement no longer threatens to supplant, but rather supplements its public counterpart. The Damages Directive has commandeered the CJEU’s promotion of national litigation and redefined its utility.

(ii) More Economic Approach

In addition to diminishing the role of national courts and the CJEU, the Commission has also sought to reassert a degree of control over the novel complex of decentralised public enforcement authorities. In order to do so, the Directorate needed to both maximise its influence over national competition authorities and limit their discretionary scope. The latter objective primarily relates to the Article 101(3) exemption procedure, a provision both particularly susceptible to the advancement of national interests and profoundly significant to policy.¹⁸⁰ The Commission has thus, in contradiction with its former practice,¹⁸¹ sought to redefine Article 101(3) as an ‘objective’ tool of economic analysis in order to marginalise the influence of national interests in the administration of Article 101(3) exemptions.¹⁸² This novel conception of Article 101(3) is generally known as the ‘more economic approach’.

Also known as the ‘effects-based’ approach,¹⁸³ this analytical framework is based on the benchmark of ‘enhancing consumer welfare and ensuring an efficient allocation of resources’.¹⁸⁴ Imported from the Chicago School of Economics, this approach to granting exemptions under Article

¹⁷⁹ Claudia Massa, ‘Sincere Cooperation and Antitrust Enforcement: Insights from the Damages and ECN+ Directives’ (2020) 16(1) *European Competition Journal* 126.

¹⁸⁰ Colomo (n 7).

¹⁸¹ Discussed below.

¹⁸² Brook (n 34) 136.

¹⁸³ See Lars-Hendrick Roeller and Oliver Stehmann, ‘The Year 2005 at DG Competition: The Trend towards a More Effects-Based Approach’ (2006) 29 *Review of Industrial Organization* 281.

¹⁸⁴ European Commission, ‘Guidelines on the Application of Article 81(3) of the Treaty’ [2004] OJ C101/97 para 33.

101(3) was introduced by guidelines accompanying Regulation 1/2003.¹⁸⁵ After first manifesting itself in the Commission's pivot to an effects-based approach in its guidelines on vertical restraints,¹⁸⁶ this strict economic philosophy was later evident in guidelines regarding horizontal agreements,¹⁸⁷ revision to the merger regulations,¹⁸⁸ and overhaul of policy regarding unilateral conduct by dominant firms.¹⁸⁹

Whilst criticisms of the Directorate's erstwhile forms-based approach certainly contributed to this conceptual pivot,¹⁹⁰ the adoption of these 'more objective standards of analysis',¹⁹¹ and the concomitant exclusion of public interest, has been necessitated by decentralisation. Whilst the prior centralised scheme of competition law demonstrated a laxity in economic analysis and the pursuit of an astounding variety of objectives, such flexibility would lead to remarkable fragmentation if devolved to national authorities. Indeed, this conceptual advancement largely offers a veneer of legitimacy and impartiality to what is effectively a reconsolidation of control on the part of the Commission. Although partially a development of principle, the 'more economic approach' is largely a necessity of pragmatism.¹⁹² The extent to which this novel conception has received NCA adherence and CJEU endorsement is discussed in Section V.

(iii) European Competition Network

Following the prioritisation of public enforcement and the promotion of an economic benchmark of analysis, the third method by which the Commission has sought to reconsolidate its power is through the organisation

¹⁸⁵ *ibid.*

¹⁸⁶ European Commission, 'Guidelines on Vertical Restraints' [2000] OJ C291/1.

¹⁸⁷ European Commission, 'Guidelines on the Applicability of Article 81 of the EC Treaty to Horizontal Cooperation Agreements' [2001] OJ C3/02.

¹⁸⁸ Council Regulation (EC) No 139/2004 on the control of concentrations between undertakings [2004] OJ L24/1.

¹⁸⁹ European Commission, 'DG Competition discussion paper on the application of Article 82 of the Treaty to exclusionary abuses' (Brussels, 2005).

¹⁹⁰ See Valentine Korah, 'From Legal Form Toward Economic Efficiency: Article 85(1) of the EEC Treaty in Contrast to US Antitrust' (1990) 35 *The Antitrust Bulletin* 1009.

¹⁹¹ Venit (n 103) 574.

¹⁹² Brenda Sufrin, 'The Evolution of Article 81(3) of the EC Treaty' (2006) 51(4) *The Antitrust Bulletin* 915, 965.

of the European Competition Network (ECN). Having established a decentralised framework of multi-level governance, the Commission has designed an infrastructure for cooperation between the NCAs in order both to develop its influence and encourage procedural uniformity. By nurturing informal conduits of influence, the Commission has bolstered its dominant position in the guidance of EU policy. Indeed, some commentators suggest that the ECN has reinstituted the Directorate as a ‘policy entrepreneur’,¹⁹³ enhancing the authority of its soft-law instruments and facilitating institutional harmonisation through osmosis. The European Competition Network represents not only an ‘innovative model of governance’,¹⁹⁴ but also an innovative model of centralisation.

Indeed, modernisation appears to be predicated on the ‘close cooperation’ of the Commission and NCAs in the application of competition law,¹⁹⁵ and the ECN was described by the Commission as a ‘core element’ of the new regime.¹⁹⁶ In addition to providing NCAs with ‘assistance’, facilitating the ‘efficient allocation of cases’, and organising the ‘mutual exchange of information’,¹⁹⁷ it was envisaged that this network would ‘promote consistency’¹⁹⁸ across the single market by encouraging the assimilation of common procedures and sanctions. By ensuring a ‘common competition culture in Europe’,¹⁹⁹ the ECN was designed to mitigate fragmentation arising from decentralisation. To a large extent,²⁰⁰ the ECN has succeeded in catalysing a degree of cultural and procedural assimilation. Commentators have observed the process of ‘institutional isomorphism’ across the Union, with the NCAs of certain Member States such as Italy, the Netherlands and Sweden being

¹⁹³ Giandomenico Majone, ‘The European Commission as Regulator’ in Giandomenico Majone (ed), *Regulating Europe* (Routledge 1996) 74.

¹⁹⁴ Alisa Sinclair, ‘Proposal for a Directive to Empower National Competition Authorities to be More Effective Enforcers (ECN+)’ (2017) 8(10) *Journal of European Competition Law and Practice* 625, 625.

¹⁹⁵ Council Regulation (EC) 1/2003 on the implementation of the rules on competition laid down in Articles 81 and 82 of the Treaty [2002] OJ L1/1, recital 15.

¹⁹⁶ Commission, ‘*Proposal for a Council Regulation on the Implementation on the Rules on Competition*’ COM (2000) 582 final 6.

¹⁹⁷ *ibid.*

¹⁹⁸ *ibid.*

¹⁹⁹ European Commission, ‘Draft Commission Notice on Cooperation within the Network of Competition Authorities’ OJ C101/43 [2003] para 3.

²⁰⁰ The ECN and procedural divergence are discussed below.

modelled on the Directorate.²⁰¹ Furthermore, the Commission is increasingly governing through soft law:²⁰² a manifestation of its 'steering activity' within the ECN.²⁰³

Some observers have analysed this system of 'dynamic delegation' through the lens of juridification, a theory concerned with the legal prescription or 'constitutionalisation' of economic systems by which economic behaviour is increasingly subject to legal rules.²⁰⁴ Accordingly, the ECN can be interpreted as transforming the relationship between the NCAs and the Commission into a 'wholly juridified relationship', in which politics is 'replaced with law supplanting power structures, decisions about competing resources, public opinion, pressure group lobbying and so on in determining the nature and content of that relationship'.²⁰⁵ Akin to the movement towards a more economic approach, the ECN is designed to minimise NCA discretion and mitigate fragmentation while consolidating the Commission's control over the complex of public competition enforcement.²⁰⁶ Indeed, by an adroit 'political masterstroke',²⁰⁷ the Commission has fashioned the decentralisation of European competition law into a Europeanisation of decentralised competition law.

²⁰¹ See Frans van Waarden and Michaela Drahos, 'Courts and (epistemic) Communities in the Convergence of Competition Policies' (2002) 9 *Journal of European Public Policy* 913.

²⁰² Despite extensive debate regarding the effectiveness of soft law (See both Linda Senden, *Soft Law in European Community Law* (Hart, 2004) 112 and Francis Snyder, 'The Effectiveness of European Community Law: Institutions, Processes, Tools and Techniques' (1993) 56(1) *Modern Law Review* 19, 32) the ECN is widely regarded as a 'successful and innovative model of governance for the complementary implementation of EU law at both European and national levels': Kathryn Wright, 'The Ambit of Judicial Competence after the EU Antitrust Damages Directive' (2016) 15 *Legal Issues of Economic Integration* 43.

²⁰³ Colomo and Kalinirini (n 12) 361.

²⁰⁴ Gunther Teubner, 'General Aspects' in Gunther Teubner (ed), *Juridification of Social Spheres* (De Gruyter 1987) 11.

²⁰⁵ Benjamin Morris, 'Juridification in the UK: The cases of the Office of Fair Trading and the Commission for Racial Equality' [2004] PhD Thesis, London School of Economics and Political Science 1, 23.

²⁰⁶ Thorough analysis of the ECN is difficult given its informal structure and lack of legal personality. The network does not issue decisions or any regulatory acts: See Ruben P Molleda, 'The ECN+ Directive and the Next Steps for Independence in Competition Law Enforcement' (2021) 12(3) *Journal of European Competition Law and Practice* 167.

²⁰⁷ Alan Riley, 'EC Antitrust Modernisation: The Commission Does Very Nicely - Thank You!' (2003) 24 *European Competition Law Review* 604, 657.

IV. FRAGMENTATION

Since the modernisation efforts that re-established, re-instituted, and redefined public intervention as the primary constituent of EU antitrust enforcement, the Commission has reiterated that a ‘substantial level of convergence in the application’ of antitrust rules has been achieved²⁰⁸ and that decentralisation has, generally, been a success.²⁰⁹ NCAs have become a ‘key pillar’ of EU competition law taking approximately 85% of enforcement decisions²¹⁰ and the Commission now principally focuses on strategically significant cases.²¹¹ Several detailed analyses of the enforcement practices of various NCAs have, however, identified significant fragmentation in the national application of EU competition law.

The foregoing analysis has framed modernisation as an attempt by the Commission to consolidate authority and reinforce public enforcement as the primary component of EU competition law in response to the jurisprudence of a Court pursuing divergent policy objectives. Although a reallocation of competence was certainly desirable to relieve the Commission of an unrealistic caseload, the total decentralisation of competition law has always been overshadowed by a fear of divergence arising from the different ‘institutional, personal and financial settings’ of each NCA.²¹² Notwithstanding the Commission’s insistence that substantial convergence has been achieved, a comparison of the enforcement practices of certain NCAs appears to confirm such concerns. Divergence arising from the deficiency of safeguards in Regulation 1/2003 against domestic legislative interference, institutional discrepancies between NCAs, and the insufficiency of soft-law guidance, raises questions regarding the prematurity of modernisation.

Additionally, persistent obstacles to convergence erected by the case law of the CJEU prove that the Commission has not eliminated the Court’s influence over EU antitrust. In addition to catalysing divergence in its

²⁰⁸ Commission, ‘Ten Years of Antitrust Enforcement under Regulation 1/2003: Achievements and Future Perspectives’ COM (2014) 453 final, 8.

²⁰⁹ Commission, ‘Report on the functioning of Regulation 1/2003’ COM (2009) 206, 3.

²¹⁰ Sinclair (n 194) 625.

²¹¹ COM (2014) 453 final (n 208) 8.

²¹² German Monopolies Commission, *Cartel policy change in the European Union? On the European Commission's White Paper of 28th April 1999* (Nomos Verlag 2000) para 40.

acceleration of modernisation, the Court has furthermore exacerbated such fragmentation through patterns of jurisprudence inconsistent with the aims of the Commission and the prior case law of the CJEU itself. This section explores divergences arising from: (i) the inconsistent application of Article 101, (ii) procedural and institutional differences between NCAs, and (iii) the disparate adoption between NCAs of the 'more economic approach' in the context of Article 101(3).

(i) Article 101(1)

Divergence in the application of Article 101(1)²¹³ has arisen for two reasons: legislative limitation and CJEU doctrine. Whilst the original legislative proposal for Regulation 1/2003 provided for the total exclusion of national law when an agreement fell within the scope of Article 101,²¹⁴ controversial reception by Member States saw the exclusion of this provision in the Modernisation White Paper.²¹⁵ National legislatures are thus permitted to limit the application of EU competition law. Whilst such legislation and the invocation thereof remains exceptional in some Member States,²¹⁶ other jurisdictions have introduced legislative limitations of Article 101 on a concerningly *ad hoc* basis. The Hungarian Parliament, for instance, has promulgated legislation exempting anti-competitive agreements in the agricultural sector in order to ensure a fair income for producers,²¹⁷ with the intention of exempting a specific anti-competitive agreement before the Hungarian NCA. This intervention led to the termination of proceedings in that case,²¹⁸ concurrent cases²¹⁹ and has seemingly deterred any subsequent

²¹³ Once again, it is to be recalled that Article 101(1) contains the blanket prohibition on anticompetitive agreements. According to the Treaties, and the CJEU's interpretation thereof (see *Société Technique Minière* (n 17)), exemption may only occur at a latter stage of analysis under Article 101(3).

²¹⁴ COM (2000) 582 final (n 196).

²¹⁵ Wouter PJ Wils, 'Ten Years of Regulation 1/2003: A Retrospective' (2013) 4(4) *Journal of European Competition Law and Practice* 293, 295.

²¹⁶ E.g. the UK Public Policy Clause examined below.

²¹⁷ Kati J. Cseres, "'Acceptable' Cartels at the Crossroads of EU Competition Law and the Common Agricultural Policy: A Legal Inquiry into the Political, Economic and Social Dimensions of (Strengthening Farmers') Bargaining Power' (2020) 65(3) *The Antitrust Bulletin* 401, 402.

²¹⁸ GVH (Hungarian NCA) [2012] *Watermelon* case (Vj-62/2012).

²¹⁹ GVH [2009] *Sugar Cartel* case (Vj-50/2009).

investigation in the agricultural sector. Similarly, the German legislature has amended the Federal Forests Act²²⁰ in response to a specific infringement finding of the German NCA regarding the sale and marketing of timber.²²¹ Such legislation leads to uncertainty in antitrust policy and has been criticised by both commentators²²² and the NCAs themselves.²²³

Secondly, the CJEU has developed doctrine to allow the exemption of the application of Article 101(1) in certain cases. Through adoption and interpretation of the ‘state action defence’ of the US Supreme Court,²²⁴ it is now established in the EU that the conduct of a Member State may be shielded from the application of Article 101 where the State engages in a *bona fide* exercise of its sovereign regulatory power.²²⁵ Furthermore, private entities may also enjoy protection under this defence where they pursue a ‘clearly articulated’ state policy and are subject to ‘active supervision’ by the state.²²⁶ Despite ruling in *INNO/ATAB* that national measures mandating anti-competitive practices would not generally be compatible with the Treaties,²²⁷ the CJEU has upheld such state measures.²²⁸ The three-step test governing this exception has, however, been repeatedly applied inconsistently by the European Courts, with the CJEU requiring cumulative satisfaction of all three criteria,²²⁹ satisfaction of selected criteria,²³⁰ and merging this three-pronged approach into a novel test on

²²⁰ Gesetz zur Erhaltung des Waldes und zur Förderung der Forstwirtschaft, Section 46(2) (German Forestry Legislation).

²²¹ Bundeskartellamt (German NCA) [2015] *Gemeinsame Rundholzvermarktung durch das Land Baden-Württemberg* (B1-72/12).

²²² Balázs Csépai, ‘The ceasefire is over’ (2015) 36(9) *European Competition Law Review* 404.

²²³ Brook and Cseres in Varju (n 5).

²²⁴ *Parker v Brown* 317 US 341 (1943) (US Supreme Court) para 5; *California Liquor Dealers v Midcal Aluminium* 445 US 97 (1980) para II, 3 (US Supreme Court).

²²⁵ See Damien Gerard, ‘EU Competition Policy after Lisbon: Time to Review the ‘State Action Doctrine?’ (2010) 1(3) *Journal of Competition Law and Practice* 202.

²²⁶ Timothy J Murris, ‘State intervention/State action—a U.S. Perspective’ in Barry E. Hawk (ed), *Fordham Corporate Law Institute* (New York, 2003) 531, 546.

²²⁷ Case 13/77 *INNO-B.M. v Association des détaillants en tabac (ATAB)* [1977] ECR 2115, para 35.

²²⁸ Case 267/86 *Pascal Van Eecke v ASPA NV* [1988] 4769.

²²⁹ See *Van Eecke* (ibid); C-2/91 *Meng* [1993] ECR I-5751; C-245/91 *Obra Schadeverzekeringen* [1993] ECR I-5851.

²³⁰ Case C-185/91 *Reiff* [1993] ECR I-5801; C-35/96 *Commission v Italy* [1998] ECR I-3851.

occasion.²³¹ Recognising the uncertainty arising from this varied application of the doctrine, academics,²³² Advocate Generals,²³³ and, remarkably, the Commission itself²³⁴ have urged the Court to clarify its jurisprudence to no avail.

Moreover, the CJEU has also established an 'inherent restrictions exception' in cases where potentially anti-competitive state measures pursue a legitimate aim. While setting out a test based on the 'objectives', 'legitimate aim', and 'proportionality' of the measures,²³⁵ the Court failed to establish 'a specific framework that could guide the assessment of these exceptions', and has facilitated the development of 'ad hoc, sector-specific' jurisprudence in national courts, leaving 'an exceptionally large margin of discretion to national competition enforcers'.²³⁶ This open-ended exception to the application of Article 101 has allowed NCAs to balance various national interests against competition law interests, giving rise to discrepancies in their application amongst NCAs,²³⁷ and between NCAs and the Commission.²³⁸

(ii) Institutional Differences

Additionally, procedural differences and varying institutional arrangements between NCAs facilitate the permeation of domestic interests and political influences into antitrust enforcement. Differences in enforcement practices between Member States arising from varying procedures governing sanctioning, investigation prioritisation, judicial review, and insulation from government influence tend to reflect national political dispositions. This can be observed through a comparison of the enforcement practices of the German Bundeskartellamt (BKartA) and the UK Competition and Markets Authority

²³¹ See Case C-94/04 *Cipolla* [2006] ECR I-11421.

²³² Gerard (n 225).

²³³ See Advocate General's Opinion in C-180/98 *Pavlov* ECLI:EU:C:2000:151, para 161: 'I do not find the case law ... very satisfactory'.

²³⁴ Reported and discussed by AG Maduro in his opinion in *Cipolla* (n 231).

²³⁵ Case C-309/99 *Wouters v Algemene Raad van de Nederlandse Orde van Advocaten* [2002] ECR I-1577, para 97. See also Case C-519/04 P *Meca-Medina v Commission* [2006] ECR I-6991.

²³⁶ Brook and Cseres (n 223) 201.

²³⁷ See Or Brook, 'Coding Non-Competition Interests Under Article 101 TFEU: A Quantitative and Qualitative Study' [2019] PhD Thesis, University of Amsterdam.

²³⁸ *ibid.*

(CMA),²³⁹ particularly in the context of the complex and rapidly emerging category of digital market infringements. It is important to note that analysis of the UK CMA is of purely historical significance given the UK's departure from the EU, and is included in the present discussion to exemplify Member State divergence. Cognisant of the harmful effects of hyper-enforcement on innovation, the CMA tends to adopt a reluctant approach to intervention in e-commerce cases,²⁴⁰ reflecting its 'liberal economy'²⁴¹ and 'neo-American' capitalist economic regime.²⁴² Conversely, the BKartA's more 'restrictive policy'²⁴³ focused on optimising social welfare has been attributed to its 'European Rhine model' of economics.²⁴⁴ The lack of procedural consistency across the NCAs of Member States permits and facilitates the permeation of such political values into competition enforcement, threatening the uniform application of competition law.

Divergence arising from such procedural incoherence is exemplified in the two NCAs' differing approaches to prioritising cases. Unlike the German BKartA in which there is no formal prioritisation procedure, the UK CMA may initiate investigation if there are 'reasonable grounds' for suspecting an infringement of competition law.²⁴⁵ Cases are chosen based on the CMA's prioritisation principles, which require a consideration of an investigation's impact on consumer welfare, strategic significance, probability of success, and demand on resources.²⁴⁶ This leads to divergent enforcement in certain e-commerce practices, such as online sales bans, which have both positive and negative effects on competition. While the BKartA's lack of a formal prioritisation procedure arguably allows it to afford undue weight to

²³⁹ An insightful analysis and comparison of the two authorities can be found in: Walker (n 4) 468. It must, however, be noted that analysis of the CMA is only of historical relevance given the United Kingdom's separation from the European Union.

²⁴⁰ Alex Schofield, 'Personalized Pricing in the Digital Era' (2019) 18(1) *Competition Law Journal* 35, 40.

²⁴¹ Christopher Townley, 'Coordinated Diversity: Revolutionary Suggestions for EU Competition Law (and for EU Law too)' (2014) 33(1) *Yearbook of European Law* 194, 243.

²⁴² See Michel Albert, *Capitalism Against Capitalism* (Whurr Publishers 1993).

²⁴³ Marc Wiggers, Robin Struijlaart and Johannes Dibbits, *Digital Competition Law in Europe: A Concise Guide* (Kluwer 2019) 128.

²⁴⁴ Schofield (n 240).

²⁴⁵ Competition Act 1998, s 25.

²⁴⁶ Competition and Markets Authority, 'Prioritisation Principles for the CMA' [2014] CMA16.

anti-competitive effects, the CMA's prioritisation principles often prevent it from investigating conduct where the detriment of which is not immediately clear.

Furthermore, the divergent mechanisms used to quantify fines can lead to inconsistent sanctioning across Member States. The CMA sets fines on an *ad hoc* basis, with a cap of 30% of an undertaking's turnover in the geographic market affected by the infringement in the past year,²⁴⁷ whilst the harsher BKartA starts fines at 10% of the infringement's total turnover throughout the course of the infringement, and further multiplies in proportion to the undertaking's total global turnover.²⁴⁸ Moreover, the BKartA's aggressive enforcement strategy has led to enormous discrepancy in fine revenue between the two authorities. Whilst the BKartA issued a total of €5 billion in fines from 2004 to 2020, the CMA imposed a mere £589 million in financial sanctions during the same period. The CMA's reluctance appears to be attributable to its fear of disincentivising innovation, particularly in the case of small undertakings, having granted significant fine reductions based on the size of firms.²⁴⁹

Differing appellate procedures and standards of judicial review additionally demonstrate the permeation of national interest through procedural inconsistencies. Prior to Brexit, it had been contended that the UK was 'the best jurisdiction in the world to defend a competition case',²⁵⁰ owing in part to a three-tier system of appeal allowing both a full merits review before the Competition Appeals Tribunal (CAT)²⁵¹ and further review on points of law before the Court of Appeal and Supreme Court,²⁵² that often led to the significant reduction of fines. Rather dramatically, the German appellate procedure not only leads to a full merits review, but rather a reinitiation of investigation by the public prosecutor.²⁵³ The fine is recalculated without reference to the original sanction and, unlike in the UK, is often increased to the

²⁴⁷ Competition and Markets Authority, 'CMA's guidance as to the appropriate amount of a penalty' [2018] CMA75.

²⁴⁸ Bundeskartellamt, 'Guidelines for the Setting of Fines in Cartel Administrative Offence Proceedings' (2013) para II 1(6).

²⁴⁹ *Ping Europe Ltd v Competition and Markets Authority* [2018] CAT 13.

²⁵⁰ UK Competition Authorities, 'The UK Competition Regime' (Report by the Comptroller and Auditor General) HC 737, Session 2015-16 (February 2016) para 2.15.

²⁵¹ See, for example, *Freeserve.com Plc v Director General of Telecommunications* [2003] CAT 5.

²⁵² Competition Appeal Tribunal Rules 2015 (SI 2015/1648), s 9(1).

²⁵³ Ordnungswidrigkeitengesetz, s 67 (Germany).

EU statutory limit.²⁵⁴ It has been observed that the tendency of the CAT to reduce fines and quash decisions of the CMA disincentivises the imposition of financial sanctions, encouraging the UK CMA to resort to alternative methods of competition enforcement.²⁵⁵ Additionally, a lower threshold governing the right of third parties to appeal in the UK²⁵⁶ further facilitates appeal and success before the CAT.²⁵⁷

The most significant factor facilitating the percolation of domestic interests into competition enforcement practices is, however, the institutional isolation afforded to NCAs. Aside from legislation, other conduits of influence can encroach upon the enforcement prerogative of NCAs. Although the UK Secretary for State may exempt agreements from infringement findings in exceptional cases,²⁵⁸ the CMA is also subject to the guidelines set out in the Department for Business, Innovation and Skills' 'strategic steer'.²⁵⁹ The German BKartA, while also subject to occasional and exceptional ministerial interference,²⁶⁰ lacks such express subordination to government policy. In practice, the CMA's subservience is manifest in its deference to governmental intervention in complex areas of competition law. On the question of the attribution of liability arising from collusive price-fixing by artificially intelligent algorithms,²⁶¹ whilst the BKartA has taken a strict approach dismissing any relief of liability,²⁶² the CMA has declined to pronounce its stance on the issue,

²⁵⁴ Fabian Badtke, Alexander Birnstiel, and Till Steinworth, 'The Cartels and Leniency Review: Germany' in John D Buretta and John Terzaken (eds), *Cartels and Leniency Review* (10th Edition, The Law Reviews, Law Business Research Ltd 2022).

²⁵⁵ See Walker (n 4).

²⁵⁶ See Competition Act, s 47(1) compared to the Gesetz gegen Wettbewerbsbeschränkungen, s 63(2).

²⁵⁷ This third party right to appeal notably resulted in a recent CMA finding in *Skyscanner Limited v Competition and Markets Authority* [2014] CAT 16 being quashed.

²⁵⁸ Competition Act, schedule 3 para 7.

²⁵⁹ See Competition and Markets Authority, 'Online platforms and Digital Advertising Market Study: Final Report' (2020) 420.

²⁶⁰ Gesetz gegen Wettbewerbsbeschränkungen, s 52.

²⁶¹ See, for example, the 'Posters and Frames' case: Competition and Markets Authority, 'Decision of the Competition and Markets Authority: Online Sales of Posters and Frames' (2016) Case 50223.

²⁶² Bundeskartellamt, 'Lufthansa tickets 25-30 per cent more expensive after Air Berlin insolvency' (*Bundeskartellamt*, 2018) <https://www.bundeskartellamt.de/SharedDocs/Meldung/EN/Pressemitteilungen/2018/29_05_2018_Lufthansa.html> accessed 11 March 2023.

highlighting that an answer ‘may be a task for government’.²⁶³ Similarly, on the issue of Big-Data, whilst the BKartA has found against Facebook’s conditions for the use and collection of data,²⁶⁴ the CMA has proposed government intervention and the introduction of a new enforcement body for such breaches.²⁶⁵

The effects of procedural divergence are moreover exacerbated by inconsistent jurisprudence from the CJEU, exemplified in case law arising from online sales bans and selective distribution schemes. The Court’s failure to conclusively reconcile an ‘object-based’ approach to such bans adopted in *Pierre Fabre*²⁶⁶ with an ‘effects-based’ approach demonstrated in *Coty*²⁶⁷ has permitted differing approaches among NCAs. The UK has seemingly adhered to the transition of the CJEU’s approach; applying the ‘infringement by object’ approach in cases prior to *Coty*,²⁶⁸ and adopting an ‘infringement by effect’ stance thereafter.²⁶⁹ The German BKartA has not, however, demonstrated such conformity, and has generally maintained an ‘object-based’ approach even following *Coty*.²⁷⁰ Whilst this lack of clarity may be addressed by the CJEU in future judgments, the confusion merely demonstrates issues posed by the procedural divergence between the practices of NCAs of the various Member States. Although the ECN has realised some degree of procedural convergence, procedural differences persist within NCAs and the broader legal systems of Member States.

²⁶³ David Currie, Chairman of Competition and Markets Authority, ‘David Currie on the role of competition in stimulating innovation’ (Speech given at the Concurrences Innovation Economics Conference, King’s College London, 3 February 2017) para 17.

²⁶⁴ Bundeskartellamt, ‘Decision on *Facebook* rendered by the Bundesgerichtshof’ [2020] BGH (KVR 69/19).

²⁶⁵ Competition and Markets Authority (n 215) para 9.7.

²⁶⁶ Case C-439/09 *Pierre Fabre Dermo-Cosmétique SAS v Président de l’Autorité de la concurrence* [2011] ECR I-9419 [41].

²⁶⁷ Case C-230/16 *Coty Germany GmbH v Parfümerie Akzente GmbH* EU:C:2017:603 [60].

²⁶⁸ Office of Fair Trading, Roma-branded mobility scooters: prohibitions on online sales and online price advertising’ [2013] (CE/9578-12).

²⁶⁹ *Ping* (n 249).

²⁷⁰ Bundesgerichtshof decision, *Asics* [2017] KVZ 41/17.

(iii) Article 101(3)

The greatest divergence in competition enforcement has, however, arisen from the decentralised application of Article 101(3) exemptions. Whilst the delegation of the Commission's monopoly over this essential antitrust provision aimed to mitigate the 'application of multiple national standards which may be different in content or enforcement from the standard of [EU] competition law',²⁷¹ the 'considerable discretionary power'²⁷² inherent in its application has inevitably led to inconsistent application across jurisdictions. NCAs differ in the extent to which they abide by the Commission's 'more economic approach' and limit the provision of 101(3) exemptions to 'direct economic benefits'.²⁷³

Some national authorities have demonstrated an adoption of this more economic approach. The UK CMA,²⁷⁴ in reference to the Commission's guidelines, has expressly declared that it shall only take account of direct economic benefits in applying Article 101(3),²⁷⁵ and has declined to examine social benefits derived from the provision of services in deprived areas²⁷⁶ and the improvement of working conditions.²⁷⁷ Similarly, the Hungarian Gazdasági Versenyhivatal has only ever accepted direct economic benefits in awarding an exemption under Article 101;²⁷⁸ holding that education, promotion of research and employment could not constitute benefits within the meaning of the Article.²⁷⁹

Conversely, other NCAs have rejected this objective economic analysis. The French Autorité de la Concurrence has maintained the applicability of

²⁷¹ Progress report from the Presidency of the Council of the European Union to COREPER, 13563/01 [10].

²⁷² European Commission (n 13) point 190.

²⁷³ Brook (n 34) 129.

²⁷⁴ It must be noted that, following the UK's departure from the EU, analysis of the UK CMA is of purely historical significance, and is included here to exemplify divergence between Member State NCAs.

²⁷⁵ Competition and Markets Authority, *Tobacco* [2016] (CE/2596-03) para 7.58.

²⁷⁶ United Kingdom Office of Fair Trading, 'Short-Form Opinion on Rural Broadband Wayleave rates' [2012].

²⁷⁷ Competition and Markets Authority, *Modeling Sector* [2016] (CE-9859/14) para 4.170.

²⁷⁸ Brook (n 34) 129.

²⁷⁹ *ibid.*

Article 101(3) to benefits such as fair trade rules, employment protection,²⁸⁰ and culture.²⁸¹ Whilst the authority has asserted that its mandate is limited to ensuring the effective operation of markets for consumers and that the promotion of other public policies was a matter for government,²⁸² it has subsequently continued to consider a wide array of benefits in applying Article 101(3) such as public health, environmental considerations, and rural development.²⁸³

These divergent approaches tend to reflect political influence. For instance, the proceduralised practice of the Dutch Autoriteit Consument to exempt agreements of beneficial consequence from a sustainability perspective,²⁸⁴ in cases such as the coordinated closure of coal power plants,²⁸⁵ collusion regarding energy-neutral renovation practices,²⁸⁶ and limitations on the sale of livestock produced in an environmentally harmful manner,²⁸⁷ has been mandated by the Minister for Economic Affairs. This followed a parliamentary request²⁸⁸ arising from demands of the Minister for Agriculture for the consideration of animal welfare and environmental benefits in competition enforcement proceedings.²⁸⁹ The NCA has, nonetheless, seemingly courted governmental intervention, and has declared that it would not ‘take action against sustainability arrangements that enjoy broad social support if all parties

²⁸⁰ Autorité de la Concurrence (French NCA) *Bleach* (05-D-03).

²⁸¹ Autorité de la Concurrence (French NCA) *Digital books* (09-A-56).

²⁸² International Competition Network, *Competition Enforcement and Consumer Welfare: Setting the Agenda* (2011) 25.

²⁸³ Autorité de la Concurrence (French NCA) *Relay Masts* (11-A-20).

²⁸⁴ Authority for Consumers and Markets (Dutch NCA), ‘Vision Document on Competition and Sustainability’ [2014].

²⁸⁵ Authority for Consumers and Markets (Dutch NCA), ‘Analysis of the Planned Agreement on Closing down Coal Power Plants from the 1980s as Part of the Social and Economic Council of the Netherlands’ [2013].

²⁸⁶ Authority for Consumers and Markets (Dutch NCA), *De Troomversnelling* [2013] (ACM/DM/2013/205913).

²⁸⁷ Authority for Consumers and Markets (Dutch NCA), *Chicken for Tomorrow* [2014] (ACM/DM/2014/206028).

²⁸⁸ Dijkgraaf/Geurts Motion of 24 January 2013, Parliamentary Documents II, 2012/2013, 33 400 XIII, no. 99.

²⁸⁹ Letter from Henk Bleker (Minister of Agriculture) to the House of Representatives re sustainability initiatives and competition law (22.11.2011).

involved such as the government, citizen representatives, and businesses are positive about the arrangements’.²⁹⁰

Whilst governmental intervention and varying competition cultures have given rise to this divergence in the application of Article 101(3), the CJEU has exacerbated this discrepancy by failing to lend juridical support to the ‘more economic approach’. When dealing with Article 101(3) exemptions, the CJEU has articulated little support for any ideological limitation of the provision to objective economic analysis. Preliminary rulings have generally elicited brief discussion as to the possibility of exemption, failing to outline the application of the conditions of the provision to the cases before the Court,²⁹¹ or merely reiterating the wording of the Article.²⁹² Judgments on appeal have generally upheld the reasoning of the Commission without further analysis²⁹³ or have focused on procedural matters.²⁹⁴

Moreover, beyond this judicial reticence, there are firmer indications of the CJEU’s disapproval of the Commission’s more economic approach. Preliminary rulings have continued to uphold the application of Article 101(3) exemptions to ‘indirect and non-economic benefits’²⁹⁵ in areas such as regulated professions,²⁹⁶ financial services,²⁹⁷ and sport in practice,²⁹⁸ and in areas such as research and development,²⁹⁹ and culture³⁰⁰ in theory. Furthermore, even after the introduction of Regulation 1/2003, the CJEU has upheld prior Commission decisions entertaining the consideration of benefits regarding financial

²⁹⁰ Authority for Consumers and Markets (Dutch NCA), ‘ACM sets basic principles for oversight of sustainability agreements’ (ACM, 2016) <<https://www.acm.nl/en/publications/publication/16726/ACM-sets-basic-principles-for-oversight-of-sustainability-arrangements2016>> [accessed 11 March 2023].

²⁹¹ Joined Cases C-403/08 and C- 429/08 *Football Association Premier League Ltd* ECLI:EU:C:2011:631.

²⁹² Case C-238/05 *AsnefEquifax* [2006] ECR I-11125.

²⁹³ Case T-193/02 *Laurent Piau v Commission* [2005] ECR II-209.

²⁹⁴ Joined Case T-259/02 to T-264/02 and T-271/02 *Raiffeisen Zentralbank Österreich AG* [2006] ECR II-5169.

²⁹⁵ *Brook* (n 34) 132.

²⁹⁶ Case C-1/12 *Ordem dos Técnicos Oficiais de Contas* [2013] ECLI:EU:C:2006:734.

²⁹⁷ *Laurent Piau* (n 293).

²⁹⁸ *Premier League* (n 291).

²⁹⁹ Case T-168/01 *GlaxoSmithKline Services v Commission* EU:T:2006:2969.

³⁰⁰ Case T-451/08 *Stim v Commission* [2013] CLI:EU:T:2013:189.

services,³⁰¹ the environment,³⁰² and sport.³⁰³ Additionally, the General Court has declared that cross-sectional clauses under the Treaties can require the consideration of non-economic benefits under Article 101(3).³⁰⁴

The failure of Regulation 1/2003 to preclude national legislative intrusion upon the application of Article 101, the failure of soft-law instruments to restrict the discretionary national application of Article 101(3), and the lack of procedural convergence among NCAs all point to the prematurity of modernisation. Whilst, as set out above, this decentralisation and the fragmentation arising therefrom may, itself, be attributed to the aggressive case law of the Court, the CJEU has further exacerbated such divergence through inconsistent and contradictory jurisprudence.

V. RECENTRALISATION

Fragmentation in EU competition law appears to have prompted a new phase in its history. In contrast to prior subtle efforts to consolidate power through adroit decentralisation, the Commission has, through the recent enactment of both the 2018 ECN+ Directive and the 2022 Digital Markets Act (DMA), suggested a novel approach: recentralisation. Both the DMA and ECN+ Directive represent recentralisation in two distinct, yet complementary, respects: ‘soft’ recentralisation and ‘hard’ recentralisation. The former regards a strengthening of the European Competition Network and greater insulation of National Competition Agencies from domestic influence. The latter arises from the resurrection of the Commission’s exclusive enforcement prerogative in the sectoral sphere of digital markets. Additionally, the DMA exhibits a new legislative trend: the adoption of a regulatory regime independent of a competition law Treaty-basis to permit centralised enforcement, allow flexible economic approach, and evade administrative constraints and judicial oversight through a subversion of meaningful standards of review. It is argued that recentralisation efforts respond to crucial flaws in modernised EU competition law.

³⁰¹ Case T-259/02 *Raiffeisen Zentralbank Österreich v Commission* [2006] ECR II-5169.

³⁰² Case T-419/03 *Altstoff Recycling Austria AG v Commission* [2011] ECR II-975.

³⁰³ *Laurent Piau* (n 293).

³⁰⁴ T-451/08, *Föreningen Svenska Tonsättares Internationella Musikbyrå v Commission* [2013] CLIEU:T:2013:189.

(i) 'Soft' Recentralisation

The foregoing analysis of the landscape of EU antitrust enforcement identifies three principal issues arising from premature decentralisation: (i) institutional and procedural divergence between NCAs, (ii) domestic political interference with NCAs' enforcement, and (iii) inconsistent adoption among NCAs of commission guidance. In light of the institutional dissonance between the Court and Commission (the history and effects of which this article has sketched), and obstacles to legislative intervention such as subsidiarity, a clear solution presented itself to the Commission: development of the ECN.

The ECN+ Directive³⁰⁵ has represented a considerable advancement in this regard. Requiring transposition by February 4th 2021, the Directive has sought to make NCAs 'more independent and effective enforcers';³⁰⁶ facilitating mutual cooperation, reaffirming and harmonising enforcement tools, and insulating NCAs from governmental interference. These provisions collectively amount to a fortification of the independence of NCAs. Whilst such independence is generally recognised as important for the good practice of competition authorities by the OECD,³⁰⁷ UNCTAD,³⁰⁸ and academics globally,³⁰⁹ it is noteworthy that the independence requirements 'that are driving NCAs further and further away from the national governments' influence are, curiously enough, counterbalanced by a broader power of the Commission'.³¹⁰

³⁰⁵ Council Directive (EC) 2019/1 of the European Parliament and of the Council of 11 December 2018 to empower the competition authorities of the Member States to be more effective enforcers and to ensure the proper functioning of the internal market (ECN+ Directive) [2018] OJ L1/14.

³⁰⁶ Adriaan Dierx and Fabienne Ilzkovitz, 'EU Competition Policy: an application of the failing forward framework' (2021) 28(1) *Journal of European Public Policy* 1630, 1643.

³⁰⁷ Enrico Alemani and others, 'New Indicators of Competition Law and Policy in 2013 for OECD and Non-OECD Countries' (2013) OECD Economics Department Working Papers No. 1104.

³⁰⁸ See UNCTAD, 'Independence and accountability of competition authorities' [2008], Note by the UNCTAD secretariat referring to other international organisations including the World Trade Organisation, World Bank, International Monetary Fund and the Organisation for Economic Cooperation and Development.

³⁰⁹ See Tay-Cheng Ma, 'Competition Authority Independence, Antitrust, Effectiveness, and Institutions' (2010) 30 *International Review of Law and Economics* 226.

³¹⁰ Molleda (n 206) 173.

The Directive provides that NCAs are to be able to perform their duties, exercise their powers, and set their priorities free from government influence.³¹¹ National authorities are to be equipped with a sufficient number of qualified staff and are to be provided with sufficient financial and technical resources to effectively enforce Article 101.³¹² Additionally, NCAs are to be granted minimum powers of investigation and sanction, including the power to inspect business premises,³¹³ to request information from undertakings,³¹⁴ issue interim measures,³¹⁵ accept commitment decisions,³¹⁶ and impose effective, proportionate, and dissuasive fines.³¹⁷ While this standardisation of enforcement capabilities, along with minimum fining amounts,³¹⁸ contributes to consistency and uniformity in the application of competition law, it also frees NCAs from reliance on national conferral of such powers. The Directive furthermore promotes closer cooperation between NCAs. In addition to imposing a general principle of cooperation,³¹⁹ it empowers NCAs to partake in investigations, receive material from, and request that their decisions be imposed in other Member States.³²⁰ The ECN+ Directive also increases the coordination of leniency programmes throughout the Member States; allowing the submission of leniency applications to multiple jurisdictions,³²¹ providing for criminal immunity to compliant personnel,³²² and requiring that leniency statements and settlement submissions enjoy the same protection at national as at EU level.

³¹¹ ECN+ Directive (n 305) Article 4. According to the ECN's Recommendation on the Power to Set priorities, less than half of NCAs had the power to set their own priorities prior to the Directive.

³¹² *ibid* Article 5.

³¹³ *ibid* Article 6.

³¹⁴ *ibid* Article 8.

³¹⁵ *ibid* Article 10.

³¹⁶ *ibid* Article 11.

³¹⁷ *ibid* Article 13.

³¹⁸ *ibid* Articles 14 and 15.

³¹⁹ *ibid* Article 27.

³²⁰ *ibid* Articles 24 and 26.

³²¹ *ibid* Article 22.

³²² *ibid* Article 23.

(ii) 'Hard' Recentralisation

Recently adopted and due to come into effect in early 2023, the Digital Markets Act seeks to ensure 'fair and open digital markets'³²³ by 'neutralis[ing] the competitive advantages enjoyed by gatekeepers'³²⁴ therein. Under this new regime, certain 'core platform service'³²⁵ providers classified as 'gatekeepers'³²⁶ shall be subject to a framework of prescriptive *per se*, *ex-ante* rules; representing a novel form of regulation and a significant 'paradigm shift' in EU competition law.³²⁷ This regulation purports to empower the Commission to identify and address structural issues in such markets at an early stage in order to prevent the entrenchment, exploitation, and leveraging of market power within gatekeepers to the detriment of third party adjacent activities and consumer welfare. Though lauded as a new 'gold standard' in the realm of digital market competition enforcement,³²⁸ the DMA also represents a stark pivot towards recentralisation in legislative policy.

In discussing the DMA, it is necessary to address its contentious interaction with EU competition law. Despite consistent academic reference to the Regulation's 'obvious proximity to EU competition law'³²⁹ and clear inspiration from antitrust cases and investigations,³³⁰ the Commission insists

³²³ Christopher Carugatti, 'The Role of National Authorities in the Digital Markets Act' (2022) 1 *Concurrences* 1.

³²⁴ Pablo Ibáñez Colomo, 'The Draft Digital Markets Act: A Legal and Institutional Analysis' (2021) 12(7) *Journal of European Competition Law and Practice* 561, 568.

³²⁵ DMA (n 1) Article 2(2) provides for the identification and definition of 'core platform services'.

³²⁶ *ibid* Article 3(1).

³²⁷ Assimakis P Komninos, 'The Digital Markets Act and Private Enforcement: Proposals for an Optimal System of Enforcement' in Eleanor M Fox, *Liber Amicorum, Antitrust Ambassador to the World* (Concurrences 2022) 425.

³²⁸ Council of the European Union, 'DSA: Council Gives Final Approval to the Protection of Users' Rights Online' (*Council of the European Union*, 4 October 2022) <<https://www.consilium.europa.eu/en/press/press-releases/2022/10/04/dsa-council-gives-final-approval-to-the-protection-of-users-rights-online/>> accessed 9 March 2023.

³²⁹ Heike Schweitzer, 'The Art to Make Gatekeeper Positions Contestable and the Challenge to Know What is Fair: A Discussion of the Digital Markets Act Proposal' (2021) 3 *ZEUP* 503, 509.

³³⁰ See Commission Staff Working Document, 'Impact Assessment Report – Annexes' SWD(2020) 363 final (Brussels, 2020).

that the DMA ‘is not a competition law instrument’,³³¹ but rather pursues ‘an objective that is complementary to, but different from, that of protecting undistorted competition on any given market’.³³² This article partially agrees. Whilst unmistakably pursuing goals analogous to Articles 101 and 102 TFEU, the DMA is, formally, an entirely independent regulatory instrument, enacted on the unorthodox Treaty basis of Article 114 TFEU.³³³

This section analyses the legislative form of the DMA. Whilst conceding that its system of predominantly centralised enforcement is appropriate in the context of addressing gatekeepers of digital markets, a more critical analysis of the Regulations suggests that the novel regime is a reaction to crucial structural, ideological, and institutional flaws of the modernised regime of EU competition law. By resurrecting a centralised enforcement prerogative, omitting reference to objective economic criteria, and eschewing meaningful standards of review, the DMA seeks to avoid fractured application, dismiss the ‘more economic approach’, and evade judicial oversight. Whilst conveniently relieving the Commission of post-modernisation encumbrances in this sector, this artful drafting suggests that traditional competition law, based on Articles 101 and 102, is not adequate to deal with complexifying markets. Whilst section V examines the practical implications of modernisation in traditional markets, the present section discusses an apparent recent legislative reaction thereto.

Both Articles 5 and 6 of the DMA address market-entry issues such as transparency in advertisement intermediation, the mobility of both end-users and business-users, unfair pricing, self-preferencing, and the interoperability of software. The former provision prohibits certain ‘black list[ed]’³³⁴ anti-competitive behaviours, whilst the latter imposes less detailed obligations upon gatekeepers, requiring further specification by the Commission and

³³¹ Commissioner Margrethe Vestager, ‘Competition in a Digital Age’ (*European Commission*, 17 March 2021)

<https://ec.europa.eu/commission/commissioners/2019-2024/vestager/announcement/s/competition-digital-age_en> accessed 9 March 2023.

³³² DMA (n 1) Recital 10.

³³³ Article 114 TFEU is the central EU treaty provision for the harmonisation of EU law in Member States.

³³⁴ See Alexandre de Streel, ‘The European proposal for a Digital Markets Act: A first assessment’ [2021] Centre on Regulation in Europe <https://cerre.eu/wp-content/uploads/2021/01/CERRE_Digital-Markets-Act_a-first-assessment_January2021.pdf> accessed 9 March 2023.

facilitating the structural redesign of concerned platforms. Whilst the prescriptive and *ex-ante* nature of these regulations represents, in and of itself, a groundbreaking deviation from prior form in EU competition law (discussed below), the DMA's conferral of exclusive enforcement authority to the European Commission represents a stark manifestation of a policy of recentralisation. Breaking from 'tradition' and suggesting 'a big step forward in European integration',³³⁵ Article 3 concentrates the power to designate 'gatekeeper' status, a prerequisite to the application of the DMA's obligations, within the exclusive (and seemingly unrestricted)^{332|336} competence of the Commission.

Evoking 'memories of the Commission's case overload ... under Regulation 17/62',³³⁷ the adoption of a system of centralised enforcement of the DMA has attracted much academic attention, with support for this legislative choice largely prevailing on account of the nature of the targeted undertakings, cost considerations, and a fear of fragmentation. Considering its focus on a remarkably small number³³⁸ of 'large, often global, firms'³³⁹ of 'pan-European reach'³⁴⁰ largely engaged in cross-border activity,³⁴¹ centralised enforcement is strategically obvious. In addition to mitigating the potential for capture by local operators and governments,³⁴² this model prevents needless duplication of enforcement efforts to address platforms that will remain similar in each Member State.³⁴³ Furthermore, centralised enforcement will avoid unnecessary coordination between NCAs and the Commission, cutting costs and preventing delays.³⁴⁴

³³⁵ Alexandre de Streel and Pierre Larouche, 'The European Digital Markets Act: A Revolution Grounded on Traditions' (2021) 12(7) *Journal of European Competition Law and Practice* 542, 558.

³³⁶ See the discussion below on administrative constraints in the DMA.

³³⁷ Schweitzer (n 329) 535.

³³⁸ The DMA's Impact Assessment suggests that there may be 10-15 such gatekeepers.

³³⁹ Alexandre De Streel and Paul Larouche, 'The European Digital Markets Act: How to Improve a Regulatory Revolution' (2021) 2 *Concurrences* 46, 58.

³⁴⁰ Impact Assessment of the DMA.

³⁴¹ Almost 25% of total online trade in Europe is cross-border. By 2025, it is expected that online marketplaces will facilitate approximately 65% of cross-border online sales in Europe: European Commission, 'Digital Markets Act: Impact Assessment Support Study – Annexes' (2020-006380) 14.

³⁴² Carugatti (n 323).

³⁴³ Monti (n 167).

³⁴⁴ *ibid.*

Converse arguments in favour of decentralised enforcement cite the capacity of NCAs to effectively receive, filter, and analyse complaints. Considering sectoral emphasis on quick intervention³⁴⁵ in conjunction with the large number of businesses that the DMA purports to protect,³⁴⁶ it has been argued that NCAs might be better placed to receive complaints from small and medium sized businesses compared to the ‘large, vertical, and bureaucratic institution’ of the Commission.³⁴⁷ Additionally, given the resources and local expertise of most NCAs, it has been argued that national enforcers are suitably positioned to act as a ‘valuable radar-screen’ or ‘primary filter’ for complaints.³⁴⁸ Similarly, given the proximity of national enforcement bodies ‘to the field’³⁴⁹ and the expectation that gatekeepers may alter various aspects of targeted platforms between Member States, NCAs have been proposed as optimal institutions to monitor compliance with the suite of obligations contained in the DMA, as has been observed in the context of EU merger control.³⁵⁰

The advantages of NCA engagement in the enforcement of the DMA may not, however, require express legislative delegation of further enforcement responsibility. Indeed, as submitted by NCAs themselves, the optimal interaction between national actors and the Commission in the enforcement of the DMA may be achieved through enhanced cooperation.³⁵¹ Interestingly, the advisory role of NCAs in the sphere of digital market ‘gatekeepers’ maintained in the DMA suggests the potential for cooperative networking to play a more pronounced role in future centralised competition regimes, highlighting the complementarity of both ‘hard’ and ‘soft’ recentralisation. Though incapable of enforcing the DMA, NCAs may issue non-binding advice on the implementation of decisions,³⁵² request the commencement of market

³⁴⁵ In a joint submission to the Commission, NCAs stressed that ‘the ability to intervene quickly is recognised as an absolute necessity in digital markets’: see European Competition Network, ‘Joint paper of the heads of the national competition authorities of the European Union: How national competition agencies can strengthen the DMA’ [2021].

³⁴⁶ *ibid.*

³⁴⁷ Mollada (n 206) 177.

³⁴⁸ European Competition Network (n 345) 7.

³⁴⁹ Larouche and de Streel (n 342) 60.

³⁵⁰ Commission Decision, *NEWSCORP/TELEPIU* (COMP/M.2876) 23.

³⁵¹ See European Competition Network (n 345).

³⁵² DMA (n 1) Article 50.

investigations through the Digital Markets Advisory Committee,³⁵³ and are involved in the Commission's adoption of Delegated Acts to add new obligations to the DMA.³⁵⁴ Prior to the adoption of the DMA, some NCAs have played a significant role in the enforcement of competition law in the digital sector; with one study attributing 48 antitrust cases, 32 merger cases, 26 market investigations, 13 position papers and 36 reports in the digital markets sector to NCA activity.³⁵⁵ Indeed, some NCAs have exhibited particular dedication to digital market infringements, with five NCAs having dedicated digital economic units, and the German Bundeskartellamt having published the first study on the intersection of competition law and data protection in the *Facebook* antitrust case.³⁵⁶ Additionally, some NCAs are particularly skilled in particular digital market issues, such as the French NCA having taken 57% of total actions by NCAs in online advertising.³⁵⁷

As EU competition legislation advances and adapts to complex issues requiring centralised and harmonised enforcement, the skills, resources, and specialisations of various NCAs may be harnessed through cooperative networks such as the ECN and Digital Markets Advisory Committee. Though the subversion of domestic influence arising from the ECN+ Directive has, in and of itself, had a 'soft' recentralising effect, it also facilitates the Commission's black-letter prerogative in the enforcement of centralised competition regulations. In addition to mitigating divergence and fragmentation in the application of Article 101 and 102 TFEU, the strengthening of the ECN may further facilitate the involvement of NCAs in the context of emerging sectors in which EU competition legislation is suggesting a centripetal tendency. It is thus contended that, given its objectives, the DMA strikes an optimal institutional balance in its enforcement regime: maintaining centralised control whilst partially engaging the expertise and resources of national actors. However, a more critical analysis of the Regulation and the underlying concerns informing its drafting suggests the emergence of a concerning legislative technique,

³⁵³ *ibid* Article 41.

³⁵⁴ *ibid* Article 49(1).

³⁵⁵ See Carugatti (n 323) 3-4.

³⁵⁶ German Bundeskartellamt, 'Bundeskartellamt Prohibits Facebook from Combining User Data from Different Sources' (*Bundeskartellamt*, 2019) <https://www.bundeskartellamt.de/SharedDocs/Meldung/EN/Pressemitteilungen/2019/07_02_2019_Facebook.html> accessed 9 March 2023.

³⁵⁷ See Carugatti (n 323) 5.

especially if the DMA is perceived to be indicative of future drafting policy in complex sectoral spheres.

Described as its principal ambition,³⁵⁸ the centralised nature of the enforcement of the DMA seeks to address a fear of ‘fragmentation’³⁵⁹ and ‘risk of divergent regulatory solutions’.³⁶⁰ By ‘internalis[ing] interstate externalities that might create a rule from a State A on a State B’,³⁶¹ the DMA seeks to avoid ‘a complex, fragmented, and possibly contradictory regulatory landscape in Europe’.³⁶² Whilst the small number and pan-European nature of the market participants targeted has principally informed the DMA’s centralised enforcement regime, the fractured enforcement of EU competition law under Articles 101 and 102 is cited as an additional justification for the exclusion of NCAs and the adoption of an independent regulatory instrument.

Moreover, the legislative peculiarity of the DMA also appears to have been informed by the inappropriate constraints of the ‘more economic approach’.³⁶³ Given the ‘multisidedness’³⁶⁴ and complexity of the business model of targeted gatekeepers, the DMA ‘dismisses the consumer welfare test’ and recalibrates the Commission’s economic approach.³⁶⁵ By prescribing *ex-ante*, *per se* obligations with variable scope, enforcement under the DMA is not qualified by reference to stringent, effects-based consumer-welfare criteria, which are arguably unsuitable to digital markets. As previously argued by Ezrachi,³⁶⁶ a broad basis of enforcement criteria is of particular importance in tackling competition infringements in digital markets. Maintaining the goal of ensuring an effective competition structure supports the pursuit of infringements ‘which are detrimental to [consumers] through their impact on an effective competition

³⁵⁸ Schweitzer (n 329) 534.

³⁵⁹ *ibid.*

³⁶⁰ See the Explanatory Memorandum to the DMA (n 3) 15.

³⁶¹ Carugatti (n 323).

³⁶² Filomena Chirico, ‘Digital Markets Act: A Regulatory Perspective’ (2021) 12(7) *Journal of European Competition Law and Practice* 493, 498.

³⁶³ Note that the ‘more economic approach’ has also been incorporated into the Commission’s practice under Article 102 TFEU, addressing abuse of a dominant market position.

³⁶⁴ Schweitzer (n 329) 536.

³⁶⁵ *ibid.* 506.

³⁶⁶ Ezrachi (n 18).

structure’,³⁶⁷ and may help enforcers to target the anticompetitive use of networks, platforms, and data tools. Maintaining the objective of protecting efficiency in the allocation of resources for the benefit of the consumer is likely to protect and foster innovation in e-commerce.³⁶⁸ Maintaining the need to protect economic freedom, plurality and democracy through competition enforcement may allow competition authorities to target search engine manipulation, ranking biases, and data harvesting;³⁶⁹ all threats to a healthy political process against which competition law is an essential weapon.³⁷⁰

In prescribing immediately applicable obligations and providing mechanisms for addition to the existing suite,³⁷¹ the DMA is likely to remedy the petrification of economic analysis brought about by modernisation by allowing the Commission to adopt an elastic economic approach in the context of digital markets. In a manner reminiscent of the early stages of EU competition law, the Commission has shed the bridle of the ‘more economic approach’ in the digital market sector by abandoning enforcement under Article 101 and fashioning a novel regime based upon Article 114. This approach sets an interesting precedent, and further highlights crucial flaws in EU competition law brought about by the phenomenon of Modernisation.

It may thus be seen that the ‘more economic approach’ and fragmentation arising from modernisation have thwarted the flexibility and utility of the traditional EU competition law regime, rendering Articles 101 and 102 inappropriate to address gatekeepers in digital markets. This has prompted the Commission to adopt an independent regulatory regime based upon Article 114, potentially indicating a novel legislative technique. The next section examines the extent to which discordance between the Court and Commission has also contributed to the legislative technique evident in the DMA.

³⁶⁷ Case 6-72 *Europemballage Corporation and Continental Can Company Inc v Commission* [1973] ECR 215, para 26.

³⁶⁸ See Alan Devlin and Bruno Peixoto, ‘Reformulating Antitrust Rules to Safeguard Societal Wealth’ (2008) 13 *Stanford Journal of Law, Business and Finance* 225.

³⁶⁹ Robert Epstein and Ronald E Robertson, ‘The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections’ (*PNAS*, 2015) <<https://www.pnas.org/doi/10.1073/pnas.1419828112>> accessed October 20th 2022.

³⁷⁰ Luigi Zingales, *A Capitalism for the People: Recapturing the Lost Genius of American Prosperity* (Basic Books 2012) 38.

³⁷¹ DMA (n 1) Article 49.

(iii) Isolation from Judicial Review

By eschewing an orthodox treaty basis for competition law, the DMA dispenses with any need for the Commission to engage with the inherent constraints of Articles 101 and 102, freeing the new regime from prior jurisprudence of the CJEU and insulating the Commission's enforcement activity from judicial review.

Articles 101 and 102 TFEU contain inherent constraints upon regulatory intervention and offer standards of review to prevent arbitrary enforcement. Infringement under the former provision demands evidence that an impugned practice has the restriction of competition as its object or effect, constraints bolstered by the case law of the CJEU. Indeed, the Court has established the necessity of case-specific legal and economic analysis in such an evaluation³⁷² and has outlined the dearth of circumstances in which practices are to be found *prima facie* unlawful irrespective of effect.³⁷³ Intervention under Article 102 requires evidence of both market dominance and the abuse of such dominance, entailing a definition of the relevant market and the examination of the features thereof in each case.³⁷⁴ These safeguards within the Treaty provisions act both to constrain and trigger the intervention of competition authorities, and offer meaningful standards to a reviewing court, delineating the parameters of what constitutes infringement under EU competition law.

Such constraints are practically absent, however, in the Digital Markets Act. Indeed, each peculiarity of the Regulation's drafting appears to subvert amenability to judicial review. Its *per se*, prescriptive 'straight-jacket'³⁷⁵ of obligations dispenses with the need or incentive for judicial clarification. Their automatic, *ex-ante* application relieves the Commission from its duty to perform specific legal and economic analysis in each case. Most significantly, the

³⁷² There is an established line of case law to this effect: See, for instance, Case C-307/18 *Generics (UK) Ltd v Competition and Markets Authority* EU:C:2020:52.

³⁷³ See Case C-228/18 *Gazdasági Versenyhivatal v Budapest Bank* [2020] EU:C:2020:265, paras 82 and 83.

³⁷⁴ On the impediments of such requirements, see Cabral and others, *The EU Digital Markets Act: A Report from a Panel of Economic Experts* (European Commission, Joint Research Centre, 2021).

³⁷⁵ Assimakis P Komninos, 'The Digital Markets Act: How Does It Compare with Competition Law?' [2022] Forthcoming in *Portale di Diritto Europeo ed Internazionale* 3.

Commission's designation of 'gatekeeper status' under Article 3, prerequisite and decisive for the application of the DMA, is subject to minimal legislative stipulation. To acquire such a classification, an undertaking must have a 'significant impact on the internal market', enjoy 'an entrenched and durable position in its operations', and provide 'an important gateway for business users to reach end users'. Though a rebuttable fulfilment of each criterion is established by reference to certain turnover and user thresholds,³⁷⁶ the Commission may alternatively designate 'gatekeeper' status to any 'core platform service' provider following a market investigation.³⁷⁷ Designation by such a market investigation merely requires the Directorate to have reference to 'some or all' of a number of elements such as the size, structure, and data-driven advantages of the undertaking. Furthermore, should the Commission fail to find such an 'entrenched and durable' position within the market by reference to such obtuse criteria, it may nonetheless designate 'gatekeeper' status if it is 'foreseeable that [the undertaking] shall enjoy such a position in the near future'.³⁷⁸ The DMA contains minimal constraints upon the application of its stringent regulations, permitting arbitrary application and undermining judicial oversight.

Moreover, unlike the Article 101(3) exception, the DMA provides no mechanism through which undertakings may be exempted from its rules on the basis of efficiency considerations or pro-competitive effects.³⁷⁹ A gatekeeper may merely submit 'sufficiently substantiated arguments' as to why the rules of the Regulation should not apply,³⁸⁰ the Commission's assessment of which is subject to no legislative stipulation. Furthermore, whereas any erstwhile competition law infringement required competition authorities to prove such a breach, the *ex-ante*, unconstrained application of the DMA instead obliges concerned undertakings to refute the application of stringent regulations, transforming and transferring a burden of proof upon the enforcer to a burden

³⁷⁶ DMA (n 1) Article 3(2).

³⁷⁷ *ibid* Article 3(8).

³⁷⁸ *ibid* Articles 3(1)(c) and 3(8)(g).

³⁷⁹ *ibid* recital 23 states that '[a]ny justification on economic grounds seeking to enter into market definition or to demonstrate efficiencies deriving from a specific type of behaviour by the undertaking providing core platform services should be discarded'.

³⁸⁰ *ibid* Articles 3(5) and 17(3).

of rebuttal upon the enforced.³⁸¹ Combined with the financial³⁸² and structural³⁸³ sanctions available to the Commission designed to penalise ‘delays in the implementation of regulatory obligations’,³⁸⁴ this reversal further disincentivises legal proceedings regarding the application of the DMA. Thus, not only does the DMA deprive the Court of its power to effectively review enforcement, it also subverts litigants’ incentive to seek review.

It must be conceded that certain sector-specific efficiency considerations have informed and partially necessitated the unorthodox form of the DMA. Limitations to the *prima facie* prohibition of vertical integration and case law, widely permitting restrictions on access to platforms,³⁸⁵ render the regime provided by Articles 101 and 102 TFEU ill-equipped to address emerging challenges posed by the entrenchment of market power within digital gatekeepers. Furthermore, the length of formal investigations have arguably necessitated recourse to directly applicable, *ex-ante* prohibitions.³⁸⁶ However, other EU competition law instruments, also drafted to adapt to technological and economic change, have not evaded judicial oversight in the same manner. Indeed, the Telecoms Directive,³⁸⁷ another prescriptive and *ex-ante* regime, subjects competition authorities to more substantial restrictions on

³⁸¹ *ibid* recital 23 states that ‘The burden of adducing evidence that the presumption deriving from the fulfilment of the quantitative thresholds should not apply should be borne by that undertaking’.

³⁸² *ibid* Articles 30 and 31.

³⁸³ *ibid* Article 25.

³⁸⁴ Colomo (n 324) 572.

³⁸⁵ As Colomo (*ibid*) highlights, a substantial number of disputes in the context of core platform service providers relate to vertical integration, a phenomenon that is not recognised by the Commission to be anti-competitive *per se*. It additionally appears that the threshold for intervention under Articles 101 and 102 is higher in the case of restrictions on access to platforms: see joined Cases C-241/91 P and C-242/91 P *Radio Telefis Éireann (RTE) and Independent Television Publications Ltd (ITP) v Commission* [1995] ECR I-743.

³⁸⁶ In Europe, a formal EU antitrust case takes an average of four years from registration to termination, with this average being protracted to seven years in many complex cases: see Digital Competition Expert Panel (lead by Jason Furman), *Unlocking Digital Competition: Report of the Digital Competition Expert Panel* (Independent Report, HM Treasury, 2019) 104.

³⁸⁷ Council Directive 2018/1972 of the European Parliament and the Council of 11 December 2018 establishing the European Electronic Communications Code [2018] OJ L321/36. A detailed comparison between this instrument and the DMA can be found in Colomo (n 324).

enforcement. The regulator may only intervene in markets not ‘effectively competitive’ where an undertaking enjoys a position of ‘significant market power’,³⁸⁸ a notion expressly defined by the CJEU in *Hoffman-La Roche* prior to the introduction of the Directive.³⁸⁹ In addition to intervention being limited to instances in which ‘significant market power is enjoyed’, the definition of which is subject to Court of Justice jurisprudence, the Telecoms Directive also subjects regulators to notions of proportionality in their intervention.³⁹⁰ Whilst the DMA evades Court influence and authority, the Telecoms regime embraces it, subjecting enforcers to limits derived from established case law, and maintaining meaningful standards of review.

Thus, it may be seen that the Digital Markets Act ‘is based on the premise that competition law principles would not limit administrative action under the new regime’.³⁹¹ Through the formulation of prescriptive, *ex-ante* provisions, the maintenance of extensive executive discretion in the designation of ‘gatekeeper status’, and the allocation of a burden of rebuttal upon targeted undertakings, the drafters of the DMA have both disincentivised and impeded judicial review of this novel regime. Indeed, by eschewing enactment upon the treaty basis of Article 101 or 102, the DMA subjects itself only to its overarching objective of ensuring ‘fairness and contestability’ in digital markets,³⁹² ‘autonomous’ notions produced ‘wholly from scratch’ under the DMA.³⁹³ Lacking any established definition, these indeterminate concepts provide little but irony, for the Regulation espouses neither. Prescriptive to gatekeepers but not to the Commission, the DMA seems hardly ‘fair’ to undertakings, and minimally ‘contestable’ before the Courts.

Though limited, the Court’s role in the enforcement and development of the Digital Markets Act has not been entirely undermined. Indeed, the Commission shall remain reliant upon the Court to ensure a degree of uniformity in effect of the DMA across Member States. The extent to which the Court complies with the objectives of the Commission’s new regime shall

³⁸⁸ *ibid* Article 67(3).

³⁸⁹ Case 85/76 *Hoffmann-La Roche & Co. AG v Commission* [1979] ECR 461, para 38.

³⁹⁰ Telecoms Directive (n 387) Article 68(2). The principle of proportionality is applicable to all Union action under Article 5(4) TFEU.

³⁹¹ Colomo (n 324) 563.

³⁹² DMA (n 1) recital 67.

³⁹³ Colomo (n 7) 590.

depend upon its interaction with and management of national courts in their dispensation of remedies to private litigants under Articles 5 and 6 of the DMA, and its preemption of existing national competition rules pertaining to digital market ‘gatekeepers’.

(iv) The Ball’s in your Court!

As an EU Regulation, the DMA shall have ‘direct effect’ in Member States, and will be ‘capable of creating individual rights which national courts must protect’.³⁹⁴ Despite its exclusion of all national actors in the designation of ‘gatekeeper’ status³⁹⁵ and textual silence on the role of private litigation in its enforcement,³⁹⁶ EU officials have stressed the ‘self-executing’ nature of the DMA,³⁹⁷ and private enforcement of regulatory obligations enshrined in Articles 5 and 6 is a certainty. Though litigation shall only provide *inter partes* relief, and shall lack an *erga omnes* effect similar to that of a decision of non-compliance by the Commission,³⁹⁸ it has been noted that national divergences in the application of available remedies, such as injunctions, specific performance, and restitutionary relief, shall lead to significant fragmentation in the enforcement of the DMA.³⁹⁹ Such divergence is likely to be compounded by the nature of the issues with which the DMA is concerned. Not only shall generalist national courts be faced with sophisticated questions of novel economic and legal complexity, but these actors shall lack any substantial body of precedence in the application thereof. Unlike the engagement of national courts under Modernisation in 2004, national judges will not be able to act in reference to 40 years of precedent developed by a specialised, centralised enforcer.⁴⁰⁰ Any divergence in the remedial practices of national courts is bound to differentiate the regulatory burden placed on gatekeepers between Member States, disturb the internal market, and diminish the effectiveness of the DMA.

³⁹⁴ Case 43/71, *Politi sas v Ministero Delle finance della Repubblica Italiana* [1971] ECR 039, para 9.

³⁹⁵ DMA (n 1) Article 3.

³⁹⁶ In a joint letter published in 2021, the Ministers for Finance of France, Germany and the Netherlands specifically called for clarity on the role of private enforcement in the context of the DMA: Bruno Le Maire, Peter Altmaier, and Monica Keijzer, ‘Strengthening the Digital Markets Act and its Enforcement’ (27 May, 2021).

³⁹⁷ Komninos (n 327) 427.

³⁹⁸ DMA (n 1) Article 29(1).

³⁹⁹ Komninos (n 327) 435.

⁴⁰⁰ *ibid.*

Furthermore, such inconsistency would be compounded by the parallel application of national regulatory frameworks in addition to the obligations imposed by the DMA. According to the principle of preemption, Article 1(5) of the DMA obliges national courts to set aside national regulations imposing 'further obligations on gatekeepers by way of laws, regulations, or administrative measures for the purpose of ensuring contestable and fair markets'. This broad prohibition is complemented, however, by a similarly obtuse exemption for national competition rules 'prohibiting anti-competitive agreements, decisions by associations of undertakings, concerted practices, [and] abuses of dominant positions'.⁴⁰¹ Whilst one might presume that the latter acts merely as an exception to the general rule of Article 1(5), this legislative dichotomy places greater responsibility upon the Court of Justice to ensure the consistent application of the DMA. Indeed, it has been observed that certain national provisions, such as Section 19 of the German Competition Act, may avoid preemption under a particular interpretation of these provisions; permitting the introduction of additional regulatory measures through the 'back door' of competition law.⁴⁰²

Thus, a familiar circumstance presents itself. Akin to the institutional dynamic arising from the Commission's consolidation of authority and subversion of judicial influence following modernisation, the Court once again finds itself as the 'ultimate safeguard'⁴⁰³ in ensuring the consistent application of EU competition law. Though some observers cite examples of the CJEU prioritising consistent and effective public enforcement to the detriment of private enforcement in other sectoral spheres,⁴⁰⁴ the pattern of interaction between the CJEU and Commission illustrated in this article suggests the Court's tendency to the contrary. Typically lenient to the autonomous inclinations of Member States and protective of the rights of private litigants, the CJEU is likely to remain disruptive to the Commission's harmonisation efforts: a thesis to be tested as it encounters issues of preemption and remedial consistency in the context of the DMA.

⁴⁰¹ DMA (n 1) Article 1(6)(a).

⁴⁰² Komninos (n 375) 6.

⁴⁰³ Komninos (n 327) 437.

⁴⁰⁴ *ibid.*

Should this Regulation be indicative of future legislative policy, it heralds a new phase of EU competition law in which institutional dynamics are dramatically augmented: a phase of recentralisation. Recent legislative developments represent recentralisation in two respects. Firstly, as a specialised, sector-specific regime, the Digital Markets Act has resurrected the Commission's prerogative in the sphere of competition law pertaining to digital markets. As an express, unequivocal reallocation of regulatory competence within the exclusive competence of the Commission, this development demonstrates 'hard recentralisation'. Secondly, and more subtly, the ECN+ Directive insulates National Competition Authorities from domestic influence, and strengthens the network connecting these NCAs and the Commission in a manner of 'soft recentralisation'. Furthermore, the DMA's exclusion of NCAs from the enforcement of its provisions is complemented by a provision for the Digital Markets Advisory Committee, allowing the Directorate to harness the expertise of NCAs through the burgeoning European Competition Network. This renders influence in the enforcement of this novel regime contingent upon active participation with and cooperation through the ECN; neatly combining the momentum and machinery of both 'hard' and 'soft' recentralisation within one mechanism.

More critical analysis may also suggest that this new phase of competition law shall be characterised by a number of evasive legislative techniques. By fashioning sector-specific regulatory regimes, based on Article 114 TFEU, the Commission may be able to escape: (i) the constraints of the 'more economic approach', a mode of analysis necessitated by modernisation, and (ii) the fragmented application of novel instruments across a Union of divergent enforcement policies. This new phase of EU competition law may also be characterised by a removal of meaningful administrative constraints. By adopting prescriptive stipulations, *ex-ante* obligations, and insubstantial doctrinal foundations, the drafters of the DMA largely subverted the function of the CJEU: recalibrating the precepts of EU competition law, further diminishing the disruptive force of the Court, but failing to obviate its influence. As the CJEU encounters issues of remedial consistency and pre-emption, the extent to which the DMA implements an effective harmonised regime shall become apparent, determining the future of the institutional relationship between the European Commission and the Court of Justice. The Commission has made its move, now the ball's with the Court.

CONCLUSION

Conflict between the policies of the Commission and the jurisprudence of the Court has spurred the development of EU antitrust, resulting in a catalysis of which the decentralisation of competition enforcement has been the primary product. Whilst necessitated by the burdensome caseload of the Directorate, the particular form and fashion of this redistribution of competence reflects the stratagem of the Commission in response to the promotion of private enforcement and the elevation of judicial influence arising therefrom.

Not only has this conflict resulted in structural reform of the Competition law system, but the Commission's subsequent modernisation attempt to consolidate and extend its control over this decentralised regime has determined its position regarding legal, economic and political matters pivotal to competition policy. The adoption of an Americanised objective economic approach, the endorsement of subsidiarity, and the cultivation of a cooperative network to organise its system of multi-level regulatory governance have all had profound effects on the discipline of antitrust. Furthermore, as problems persist with the uniform application of competition law, the Commission has even deviated from the very foundations of EU antitrust in an effort to avoid disruptive judicial oversight: eschewing orthodox treaty bases and adopting unusual legislative tact in recent enactments. Indeed, the effects of this conflict pervade all elements of competition law. This article thus concludes on the ironic deduction that competition law is, itself, shaped by competition: not a competition between undertakings, but a competition between the Court of Justice and European Commission.

A contemporaneous evaluation of the degree of success
of the Financial Markets Test Case Scheme in the seminal
case of *Financial Conduct Authority v Arch Insurance (UK)
Ltd and Others* (2021) UKSC 1

Sean Burke*

ABSTRACT

*COVID-19 had catastrophic consequences on all parts of day-to-day life, especially the economy. Businesses were mandated to shut down on public health grounds, costing them huge sums of money and destroying some altogether. Many businesses that did manage to survive found themselves in dire situations and subsequently instigated business interruption insurance claims (insurance which covers loss of income arising from some event or disaster, such as COVID), for an indemnity to cover the money which had been lost. Such an indemnity operates to ensure businesses can survive through otherwise catastrophic disasters and situations, with different policies covering differing heads of loss. The sheer scale of this public health emergency resulted in an extraordinary number of claims being submitted. Unsurprisingly, these claims were largely rejected by insurers. This left hundreds of thousands of policyholders in uncertain positions as, in many cases, legitimate claims were rejected. The Financial Conduct Authority recognised the sheer scale of the potential fallout of a lack of insurance to businesses and thus sought to commence a special type of claim under the Financial Markets Test Case Scheme under the newly founded Financial List. This Scheme was conceived to provide a forum for the expedited resolution of financial market issues of general importance as a whole. This article will evaluate and assess the degree of success of the use of this Scheme in the case of *Financial Conduct Authority v Arch Insurance (UK) Ltd and Others*.*

* LLB (Winton), LLM (Soton), LLM (London). The Honourable Society of the Inner Temple. Views expressed are solely the author's own.

INTRODUCTION

At the time of writing, the COVID-19 pandemic has largely receded, lockdown seems a memory of the past, and no businesses remain closed due to Government restrictions. This was not always the case. The emergence of COVID-19 posed such a significant health risk that the Government had to impose draconian restrictions, forcing many businesses to close. These restrictions cost businesses huge sums of money, and some were subsequently forced to close permanently. Many surviving businesses found themselves in an appalling financial condition and, consequently, had to initiate business interruption insurance claims to recover an indemnity that reflected the losses caused by these restrictions. The scale of the pandemic led to an enormous number of such claims being brought forward. There was inconsistency between insurers in accepting and rejecting claims, with many legitimate and proper claims being rejected, causing substantial confusion for hundreds of thousands of policyholders. Thus, the Financial Conduct Authority ('FCA') had to acknowledge the potential for serious negative economic consequences and they initiated a specialised type of proceeding in the (at that time) new Financial List.¹

The Financial List was a significant step forward and was initiated because of its axiomatic importance in the financial services sector in the UK, which contributed £132 billion to the UK economy in 2019 and generated 6.9% of total economic output.² The importance of this sector was instrumental in

¹ 'Result Of FCA's Business Interruption Test Case' (FCA, 2015) <<https://www.fca.org.uk/news/press-releases/result-fca-business-interruption-test-case>> accessed 3 July 2022; Bank of England, 'The Impact Of Covid-19 On Businesses' Expectations: Evidence From 'The Decision Maker Panel' (*Bank of England*, 2020) <<https://www.bankofengland.co.uk/quarterly-bulletin/2020/2020-q3/the-impact-of-covid-19-on-businesses-expectations-evidence-from-the-decision-maker-panel>> accessed 1 August 2022.

² Georgina Hutton, Ali Salchi and Matthew Ward, 'Financial Services: Contribution To The UK Economy', (*House of Commons Library*, 1 September 2022) <<https://commonslibrary.parliament.uk/research-briefings/sn06193/>> accessed 26 February 2022; D. Reid, 'New York Stretches Lead Over London As The World's Top Financial Centre, Survey Shows' (*CNBC*, 19 September 2019) <<https://www.cnn.com/2019/09/19/new-york-beats-london-again-as-the-worlds-top-financial-center.html>> accessed 7 June 2021.

the creation of the Financial List, which was first proposed by Lord Thomas in his capacity as Lord Chief Justice in July 2014. He stated that the ‘prosperity of the nation’ was dependent on the Courts providing ‘fast, efficient and economical dispute resolution’.³ This was implemented in the ensuing year, culminating in the Dinner for Her Majesty’s Judges speech, by the same Lord Chief Justice, in July 2015. In this speech, it was announced that ‘we [the Ministry of Justice] are now ready to introduce a Financial List’.⁴ The reasons given for the introduction of this Financial List expressly denoted in the speech were, *inter alia*, that:

First, it will promote access to the courts and the expertise of trial judges, for market actors in an area that is of significant importance to the development of both the domestic economy, and to open markets internationally. Secondly, and particularly through the test case procedure, the Financial List will help to avoid costly and time-consuming litigation, through providing a mechanism for authoritative guidance before disputes have arisen. Thirdly – and flowing from the first and second points – I hope that this initiative will promote the rule of law both nationally and internationally.⁵

It is the second of these reasons which are of particular importance and relevance to this article. The ‘test case procedure’, more properly called the Financial Markets Test Case Scheme (hereinafter ‘the Scheme’), was created to allow issues which have been categorised as being of ‘general importance’ to the wellbeing of the financial markets as a whole, and which require ‘immediate relevant authoritative English law guidance’, to be heard in the Financial List without the usual requirement of a present cause of action or dispute between

³ Lord Chief Justice of England and Wales Thomas of Cwmgiedd, ‘Speech’ (Dinner For Her Majesty’s Judges, London, 9 July 2014) <<https://www.judiciary.uk/wp-content/uploads/2014/07/lcj-speech-mansion-house-dinner-for-hm-judges.pdf>> accessed 26 February 2023, para 13.

⁴ Lord Chief Justice of England and Wales Thomas of Cwmgiedd, ‘Speech’ (Dinner For Her Majesty’s Judges, London, 8 July 2015) <<https://www.judiciary.uk/wp-content/uploads/2015/07/speech-lcj-mansion-house-2015.pdf>> accessed 26 February 2023, 3.

⁵ Lord Chief Justice of England and Wales Thomas of Cwmgiedd (n 4) 4.

the parties. Claims can be brought by way of mutual agreement between market participants with opposing interests in an issue which is to be resolved. The Court must be satisfied that the arguments of all parties with opposing interests will be properly represented before the court, and for that purpose may allow relevant trade, professional or regulatory bodies, or any affected third parties, to be joined as a party or otherwise represented in the proceedings.⁶ It is these characteristics of the Scheme that seemingly made it highly suitable to hear the intricate and complicated case of *Financial Conduct Authority v Arch Insurance (UK) Ltd and Others*⁷ (hereinafter 'FCA test case').

On 9th June 2020, the FCA commenced one of these test cases against eight defendant insurers, as well as Lloyd's managing agents. The case addressed issues of legal principle raised by COVID-19 business interruption claims made under coverage clauses, which are clauses defining the scope of cover that expressly do not require actual property damage. At the preliminary case management conference, which was held on the 16th of June 2020, Mr Justice Butcher decided that the case raised issues of general importance to the financial markets because they were relevant to widely-used policy wordings. Mr Justice Butcher ruled that the Scheme was applicable to the claim, and gave directions for an expedited trial before him and Flaux LJ to commence on 20th July 2020. This was the first time the Scheme had been utilised, despite being introduced several years earlier.⁸ The case progressed as follows:

On 1 May 2020, the FCA issued a statement setting out their intention to obtain court declarations aimed at resolving the contractual uncertainty around the validity of many business interruption claims. On 15 May 2020, the FCA invited policyholders of business interruption insurance, who were in dispute with their insurers over the terms of their policies, to submit their arguments to the Court by 20 May 2020 if they wanted the FCA to take them

⁶ 'Practice Direction 63AA PD 6.5.A.– Financial List' (*Justice*, 2020)

<<https://www.justice.gov.uk/courts/procedure-rules/civil/rules/financial-list/practice-direction-63aa-financial-list>> accessed January 31, 2023.

⁷ *Financial Conduct Authority v Arch Insurance (UK) Ltd and Others* (2021) UKSC 1, (2021) 2 WLR 123.

⁸ 'Test Case Concerning COVID-19 Business Interruption Claims Becomes The First Case To Proceed Under The Financial Markets Test Case Scheme' (*Fountain Court*, 2020) <<https://www.fountaincourt.co.uk/test-case-concerning-COVID-19-business-interruption-claims-becomes-the-first-case-to-proceed-under-the-financial-markets-test-case-scheme/>> accessed 4 July 2021.

into account as part of the test case. As noted above, on 16 June 2020 the first Case Management Conference commenced before Mr Justice Butcher. Mr Justice Butcher considered the application for expedition and admission to the Financial Markets Test Case Scheme, fixed the timetable for the case (including the date for the court hearing in the second half of July) and dealt with other procedural matters. On 10 June 2020, proceedings commenced in the High Court. At the first hearing on 30 June 2020, Lord Justice Flaux and Mr Justice Butcher gave various directions relating to how the test case was to proceed. On 13 July 2020, the FCA's and Interveners' skeleton arguments (written submissions) were published for a trial commencing on 20 July. On 15 July 2020, the Defendants' skeleton arguments (written submissions) were published for trial.

The trial commenced on 20 July 2020 and concluded on 31 July 2020. On 15 September 2020, the FCA published the judgment in the test case, which was handed down by the High Court. On 29 September 2020, the deadline for parties to apply for a 'leapfrog' appeal to the Supreme Court passed. The Supreme Court published a press release stating the FCA's position on the appeals process. On 16 November 2020, the parties agreed to the full running order for the Supreme Court hearing. On 15 January 2021, the Supreme Court handed down its judgment, substantially allowing the FCA's appeal and largely dismissing the insurers' appeals. On 3 March 2021, the finalised guidance and feedback statement was provided. On 14 July 2021, the FCA published the Supreme Court Order varying the High Court declarations made in the test case.

This article will evaluate and assess the degree of success of the use of the Scheme in this FCA test case. This analysis will proceed in three sections. Each section will attempt to answer the various research questions stated below, and the answers to these questions will provide a basis for evaluating the degree of success of the Scheme in the FCA test case. The first section, entitled 'The Scheme', will explore the Scheme in depth. It will examine what the Scheme is, why the Scheme was utilised, how a claim can qualify for the Scheme, and finally the purpose behind the use of the Scheme in the FCA test case specifically. The consideration of these questions will allow an assessment to be formed as to the suitability of the Scheme to hear the FCA test case. The objectives of the Scheme must align with the objectives of the FCA test case — as they both had an overriding necessity for speed and effectiveness — and so it will be assessed

whether this necessity was satisfied. This will aid determination of whether the Scheme was successful in the FCA test case in the ensuing chapters.

The second section, entitled ‘The Disputes’ will evaluate whether the disputes which were considered were satisfactorily resolved. The criteria for evaluating this question will be derived from the objectives of the Scheme and the objectives of the FCA test case as a whole. Thus, for each dispute considered, it will first be considered whether immediately relevant and authoritative English law guidance was provided, per the objectives of the Scheme. Once this has been determined, it will subsequently be considered whether this guidance clarified the relevant key issues of contractual uncertainty, per the objective of the FCA test case. The first dispute relates to disease clauses, followed by disputes relating to prevention of access and hybrid clauses. The second dispute relates to causation. This is followed by evaluations of disputes relating to trends clauses, then finally an evaluation of disputes relating to the controversial *Orient Express Hotels*⁹ case. The FCA test case can only be considered successful if these disputes are resolved in a satisfactory manner. The resolution of these disputes must be conducted in accordance with the objectives of the process utilised to resolve them. In essence, unless it is concluded that the resolution of these disputes provides immediate relevant authoritative English law guidance which clarifies key issues of contractual uncertainty, then the use of the Scheme in the FCA test case cannot be considered successful as its overriding objective would not have been met.

The third and final section, entitled ‘The Success’, follows from the evaluation of the success of the resolution of the disputes in the FCA test case, and will evaluate the role of the Scheme in coming to these resolutions. This section will explore whether or not the Scheme resolved the issues mentioned in the second section, and further whether the resolution of these issues has generated any new or unforeseen issues. It will evaluate whether the Scheme achieved its overall purpose in the FCA test case. This section will additionally evaluate whether the Scheme has improved financial services dispute management and whether the pre-existing Financial Ombudsman Service would have produced a better outcome. This discussion will progress into evaluating whether the Scheme should be extended to cater to other areas of law and

⁹ *Orient-Express Hotels Ltd v Assicurazioni Generali SpA* (UK) [2010] EWHC 1186 (Comm), [2010] Lloyd's Rep IR 531.

whether its permanence in the Financial List is a positive development. The fundamental objective of this article is to undertake a contemporaneous evaluation of the degree of success of the Scheme in the seminal FCA test case. The discussion within this article will assist in making such an evaluation.

CHAPTER I – THE SCHEME

(I) WHY UTILISE THE SCHEME?

The first part of this section will explore why the Scheme was utilised in the FCA test case to begin with. It will look at the intrinsic necessity for an effective outcome in an unusually short period of time, and further whether the Scheme would be an appropriate means of achieving this. The FCA was entirely aware of the degree of complexity which characterised the issues surrounding the business interruption policies (as they were designed to cover the loss to businesses during periods of interruption to their revenue, as in the present case) and the extent to which these issues had the potential to cause a lot of trepidation for both insurers and policyholders alike.¹⁰ For the benefit of both, it was recognised that there was a need for a quick and clear elucidation of the matters to determine which policies existed to provide cover. This was largely because the FCA acknowledged that there was such a wide range of policy wordings and resultant types of insurance covers that it was difficult to ascertain whether or not their cover included the relevant COVID-19 business interruption claims.¹¹ The genuine doubts surrounding these policies and their proper interpretation led to uncertainty, which eventually culminated in many feasible claims being denied by the insurers at what was, for some, their darkest time.¹²

¹⁰ *Financial Conduct Authority* UKSC (n 7) [3].

¹¹ Sarah Turpin and Alexander Bradley-Sitch, 'COVID-19: Considering The Wider Impact Of The FCA's Test Case' (*K&L Gates Hub*, September 2020) <<https://www.klgates.com/COVID-19-Considering-the-Wider-Implications-of-the-FCA-s-Test-Case-9-22-2020>> accessed 22 July 2021.

¹² 'Civil Justice Statistics Quarterly' (*Gon.uk*, 2021) <<https://www.gov.uk/government/collections/civil-justice-statistics-quarterly#2019>> accessed 21 July 2021; Pinsent Masons, 'An overview of civil proceedings in England and Wales' (*Pinsent Masons*, 2011) <<https://www.pinsentmasons.com/out-law/guides/an-overview-of-civil-proceedings-in-england-and-wales>> accessed 21 July 2021.

The amalgamation of these factors and circumstances compounded by the COVID-19 pandemic led to a situation whereby any uncertainty required amelioration within a short timeframe. The issue with normal litigation, however, is that it is notoriously time-consuming and requires an actual dispute or cause of action between two parties.¹³ This rendered the usual course of litigation futile for the proposed purpose of the action, which was to reach a clear and expedited decision. This led to the Scheme being proposed, and not long after, being quickly accepted, as it was recognised that it could potentially be the primary means by which clarity could be provided in a timely manner.¹⁴ The FCA subsequently commenced proceedings in pursuit of a declaratory judgment with regard to the interpretation of numerous representative sample business interruption policy wordings.

The FCA explained that the proposed test case needed to explore a variety of sample representative business interruption policies and was consequently compatible with an expedited court process, such as the process ingrained within the Scheme. Moreover, the test case was not intended to resolve all of the disputes nor provide a comprehensive analysis with regard to the multitude of situations in which policyholders found themselves, but rather to quickly provide clarity and authority to some of the most pertinent contractual uncertainties and key causation issues.¹⁵ For example, the FCA test case did not endeavour to determine precisely how much is payable under each policy wording as would likely be the case in individual litigation, but instead provided the means for doing so through the analysis of the operation of the various clauses. The FCA commenced this action in the public interest, obtaining guidance and authority in the way it did by means of the Scheme, and allowing for a decision to be provided in a faster and cheaper way than through litigation.¹⁶

¹³ 'Civil Justice Statistics Quarterly' (*Gon.uk*, 2021) <<https://www.gov.uk/government/collections/civil-justice-statistics-quarterly#2019>> accessed 21 July 2021.

¹⁴ *Financial Conduct Authority v Arch Insurance (UK) Ltd and Others* [2020] EWHC 2448 (Comm).

¹⁵ Sarah Turpin, Oliver Richardson and Edward Brown-Humes, 'Covid-19: FCA Business Interruption Test Case - Unresolved Issue And Wider Implications Of Supreme Court Judgment' (2021) 11 (75) *The National Law Review* 210.

¹⁶ *ibid.*

(II) A 'QUALIFYING CLAIM' UNDER THE SCHEME

This section will look at precisely what constitutes a qualifying claim under the Scheme. It will look at the relevant authoritative guidance and will determine whether the FCA test case fell neatly under the criteria for a claim to be classified as a qualifying claim, or alternatively, whether the parameters of the criteria were stretched in order for the case to be heard under the Scheme. This is important because, as previously mentioned, the FCA test case was the very first case to be considered under the Scheme, despite it being introduced several years prior.¹⁷

Like many of the rules and procedures relating to civil litigation, the Scheme is encompassed in a Practice Direction: namely Practice Rules (CPR) Part 63AA. At paragraph 6.1, the Practice Direction denotes that a 'qualifying claim' for the Scheme must be 'a Financial List claim which raises issues of general importance in relation to which immediate relevant authoritative English law guidance is needed'.¹⁸ This inclusion of the Financial List at the beginning of the requisite qualifications for the Scheme is particularly important. The Financial List Guide, which is found within Practice Direction CPR Part 63A of the White Book, acknowledges that parties are permitted to transfer an action into the Financial List under the protection of the CPR Part 30, even in cases which fail to conform to the specific criteria of the Financial List.¹⁹ This allows cases which fall 'within the spirit but not the letter', of the stipulated criteria, admission into the Financial List.²⁰ This is especially important because the Financial List Guide further states that:

¹⁷ 'Test Case Concerning COVID-19 Business Interruption Claims Becomes The First Case To Proceed Under The Financial Markets Test Case Scheme | Fountain Court Chambers' (*Fountain Court Chambers*, 2021) <<https://www.fountaincourt.co.uk/test-case-concerning-COVID-19-business-interruption-claims-becomes-the-first-case-to-proceed-under-the-financial-markets-test-case-scheme/>> accessed 4 July 2021.

¹⁸ 'Practice Direction 63AA - Financial List' (*Justice.gov.uk*, 2020) at 6.1 <<https://www.justice.gov.uk/courts/procedure-rules/civil/rules/financial-list/practice-direction-63aa-financial-list#fin>> accessed 5 July 2021.

¹⁹ UK Government, 'The White Book 2022 (Volume 2)' (*Sweet and Maxwell* 2017), 942-943.

²⁰ Ministry of Justice, 'Guide To The Financial List' (*Ministry of Justice*, 1 October 2015) <<https://www.judiciary.uk/wp-content/uploads/2022/11/financial-list-guide.pdf>> accessed 11 March 2023.

case[s] concerning insurance, re-insurance or professional negligence, or a case falling within the normal specialist jurisdiction of the Companies Court (insolvencies, capital reductions, schemes of arrangement as well as shareholder disputes like unfair prejudice petitions and equitable petitions) will not generally fall within the definition of Financial List Claims. However, if issues arising in such a case were to require financial market expertise or were issues of general market importance, then it may be appropriate to issue the claim in the Financial List or transfer such a case or part of it into the Financial List.²¹

Despite the fact that the FCA test case provided guidance on insurance, which per the above quotation would preclude it from being admitted onto the Scheme, it was permitted into the Financial List and onto the Scheme because of the final part of the preceding statement, namely that the case required financial market expertise and involved issues of general market importance to English law.²² In particular, it assessed legal principles in relation to policy coverage under a number of different specimen policy wordings which had been underwritten by the defendant insurers with respect to claims by policyholders to be indemnified for business interruption losses arising from the COVID-19 pandemic, and the consequential measures which had been imposed by the UK Government. Thus, the scope of the claims within the FCA test case fell within the Financial List and, subsequently, the concerned parties were able to utilise the Scheme.²³

In the court of first instance, Mr Justice Butcher and Flaux LJ were tasked with determining the correct construction of the policy terms and whether cover was to be available in principle, with reference to a set of mutually-agreed and assumed facts between the parties in accordance with the

²¹ *ibid* para 2.3.

²² Sabur Malik, 'The Empire Strikes Back: Derivatives Disputes, The Financial List, Test Case Scheme And Arbitration' [2021] *Capital Markets Law Journal*.

²³ John Jarvis QC, Dr Stuart Dutton and Michael Green, 'Financial List In English High Court Opens Its Doors To Cases' [2021] *Journal of Law and Medicine; Financial Conduct Authority UKSC(n 7)* [2].

stipulated Framework Agreement.²⁴ The fairly unusual notion of having an even number of judges is a facet of the Scheme²⁵ that operated to only hinder any possibility of swift progression of the FCA test case. In the normal course of events, a dispute is decided in line with the majority of the Bench. To always establish a majority, there must be an odd number of judges. This is a simple and well-recognised principle of the English legal system.²⁶ Thus, the obvious potential problem with having two judges is that in the event of a disagreement, the Scheme's advancement may come to a standstill: both judges' rulings carry exactly the same weight, and it becomes an impossible task to determine a fair outcome. There is no logical reason as to why the Scheme was designed in this manner, and there is a complete omission of any guidance on what to do in such an impasse. There exists some authority on the matter, however, found embedded within s 54 and s 66 of the Senior Courts Act 1981. This guidance relates specifically to matters being heard in the High Court of England and Wales and subsequently would *prima facie* be applicable to the FCA test case at first instance.²⁷

Section 54 of the Senior Courts Act 1981 expressly denotes that, in the scenario where two High Court judges consider an appeal from a lower court and they offer differing views on the outcome of the case, the appeal should stand. This is a process which was simply not suitable for the FCA test case. It was a claim that commenced in the High Court and so there was no prior judgement following which a decision could be made. The whole point of the FCA test case being admitted onto the Scheme was to create authoritative precedent and to do so quickly.²⁸ This accompanying statutory commentary additionally explains that, wherever the situation makes it viable, the provisions relating to disagreement between members of the Court of Appeal will be applied, by analogy, to disagreement between two judges in a Divisional Court.²⁹ The practice in the Court of Appeal where an even number of judges are equally divided on the outcome is that the case shall, on the application of any party to the appeal, be re-argued before and determined by an uneven number of judges

²⁴ *Financial Conduct Authority* EWHC (n 14) [1]-[2].

²⁵ *ibid* [2].

²⁶ Alisdair A Gillespie and Siobhan Weare, '14. Those in Court' in Alisdair A Gillespie and Siobhan Weare, *The English Legal System* (Oxford University Press 2021) 500-504.

²⁷ Senior Courts Act 1981, s 54, 66.

²⁸ *ibid*.

²⁹ *ibid*.

not less than three, before any appeal to the Supreme Court. Therefore, it may be that in Financial List cases where the two presiding judges disagree, a party may apply for the test case to be re-argued before three judges before any appeal will be determined.

This is one of the factors behind ‘leapfrogging’ the appeal straight from the High Court up to the Supreme Court.³⁰ Throughout the preliminary stages, as well as throughout the High Court case, there was a particular emphasis on speed. In fact, one of the factors that qualified the FCA test case for the Scheme was the necessity for a quick and efficient outcome. At paragraph 9.1, the Financial List Guide expressly states that the procedure’s object is to provide a process through which ‘immediate relevant authoritative English law guidance’ can be sought.³¹ Appeals can be long-winded and would serve only to postpone the provision of any actionable guidance on incredibly serious matters. This is another advantage of the Scheme. It is flexible to the needs of the people who have an interest in the decision which is being sought. Claims of this nature, which require quick conclusions on practical financial market matters, are to benefit enormously from this flexibility.³²

This flexibility is found again in the Practice Direction CPR Part 63A, at paragraph 6.3. It is additionally stated that for a claim to be classified as a qualifying claim, it must be

advanced by a person who is or was actively involved in business in the relevant market against another person who similarly is or was actively involved in the business in the relevant market, provided that other person has opposing interests as to how the law of England and Wales issue(s) raised by the qualifying claim should be resolved.³³

³⁰ Katie Scott, ‘FCA Files “Leapfrog” Appeal Application To The Supreme Court For BI Test Case’ (*Insurance Times*, 2020) <<https://www.insurancetimes.co.uk/news/fca-files-leapfrog-appeal-application-to-the-supreme-court-for-bi-test-case/1434761.article>> accessed 24 July 2021.

³¹ ‘Guide to the Financial List’ (*Judiciary.uk*, 2015) <<https://www.judiciary.uk/wp-content/uploads/2022/11/financial-list-guide.pdf>> accessed 26 February 2023.

³² Malik (n 22).

³³ Practice Direction (n 18) para 6.3.

In the FCA test case, however, the FCA could not accurately be described as ‘a person who is actively involved in business in the relevant market’. The FCA is undoubtedly involved in the relevant market. In the words of the Association of British Insurers, the FCA possesses and exercises enforcement jurisdiction with regard to regulating how insurance firms behave, as well as more broadly upholding the integrity of the UK insurance market.³⁴ Yet, it would simply not be accurate to state that the FCA is involved in any business in the relevant market: it may be involved in the relevant market while not conducting business there.

Notwithstanding this factual barrier, the claim was still permitted for determination through the Scheme. This is because the FCA was representing policyholders who purchased relevant business interruption policies, and these people could be accurately described as people who were actively involved in business in the relevant market. This fairly lenient approach to qualification demonstrates that the court recognised both the importance of the matters at hand as well as the suitability of the Scheme to determine such disputes.³⁵ Thus, the FCA was able to represent the interests of the insurance policyholders, the vast majority of which were small or medium-sized businesses (SMEs) in dire need of clarity and support. Once the claim had been categorised as a ‘qualifying claim,’ it could then proceed through the preliminary stages of the case. This was done at the first case management conference, and in the FCA test case was completed by Mr Justice Butcher.

(III) THE PURPOSE OF THE SCHEME IN THE FCA TEST CASE

The third and final part of this chapter will look at the purpose of the Scheme in the FCA test case specifically. This is perhaps the most important consideration for the purposes of this article. It is essential that the purposes behind the introduction of the Scheme align with the requirements of the FCA test case (or any test case for that matter). If the two are explicitly unaligned, then it cannot be said that the Scheme was a success in the FCA test case.

³⁴ Association of British Insurers, ‘How our industry is regulated’ (*ABI*, 2023) <<https://www.abi.org.uk/data-and-resources/tools-and-resources/regulation/>> accessed 10 March 2023.

³⁵ Nicholas Bartlett and Chris Ives, ‘FCA Business Interruption Test Case: Cover For COVID-19’ (2020) 30 (10) *PLC Magazine* 4-5.

Whether the aims and purposes of the Scheme and the FCA test case align assists in determining whether utilising the Scheme for the first time was in fact the optimal means of resolving the disputes at hand. The FCA test case had an overriding necessity for speed and effectiveness, the satisfaction of which will be assessed. The degree of success will be evaluated according to the satisfaction of the objectives of the Scheme and the FCA test case, and these objectives will be explored to assist in this evaluation. Consideration and evaluation of these three aspects of the Scheme will later assist in determining whether the Scheme has been a success in the FCA test case.

As stated earlier, the Scheme was utilised in the FCA test case due to its potential to resolve important issues in short periods of time. In the normal course of things, a case such as the FCA test case would take many months to reach the High Court, and many years to reach the Supreme Court.³⁶ The Scheme provides a means by which a court can grant declaratory relief in an action where there is no ‘real’ dispute, which is the reason the claim in the FCA test case has been termed friendly action. This is another fairly unique aspect of the Scheme, the success of which shall be judged in the ensuing sections. The ability to grant declaratory relief in a friendly action is qualified, per the ruling in *Rolls-Royce PLC v Unite the Union*.³⁷ In short, the action must be brought in the public interest; there must be a pre-agreed set of facts to establish the context in which the decision is to be made; and the court must be satisfied that all parties will have an opportunity to properly advance their case.

More specifically, the primary purpose of the utilisation of the Scheme, whether in the FCA test case or in any test case at all, is to provide immediately relevant authoritative English law guidance. This has been expressly stated in the preliminary stages of the introduction of the Scheme, as well as in the preliminary stages of the FCA test case.³⁸ Yet, this purpose must operate to fulfil the purpose of the FCA test case. The overarching reason why the FCA test

³⁶ *Financial Conduct Authority* UKSC (n 7); ‘Civil Justice Statistics Quarterly’ (*Gon.uk*, 2021) <<https://www.gov.uk/government/collections/civil-justice-statistics-quarterly#2019>> accessed 21 July 2021.

³⁷ *Rolls-Royce PLC v Unite the Union* [2009] EWCA Civ 387.

³⁸ *Financial Conduct Authority* UKSC (n 7); ‘Test Case Reassures On COVID Insurance’ (*New Law Journal*, 2021) <<https://www.newlawjournal.co.uk/content/test-case-reassures-on-covid-insurance>> accessed 26 February 2023.

case was initiated was to provide guidance to clarify key issues of contractual uncertainty. This reason is the standard against which success can be measured. The resolution of the disputes must subsequently not only offer guidance on the matters at hand but must further ensure that any such guidance resolves the contractual uncertainty that the FCA test case intended to resolve.

The three preceding sections evaluated the suitability of the FCA test case in being admitted to the Scheme, and further highlights the advantages and disadvantages of utilising the Scheme in the FCA test case. The first section explored why the Scheme was used in the first place. It found that the Scheme was used because of its hugely expedited process and its subsequent ability to come to a decision in an unusually short period of time, despite having to tackle complex and intricate issues.

The second section looked at a qualifying claim under the Scheme, and whether the FCA test case falls within the parameters of the requisite criteria. It illuminated the flexibility of the Scheme and how this has massively assisted the fast resolution of the FCA test case by, for example, taking a lenient approach to the inclusion of the FCA.

This led to the third section, which examined the purposes behind the Scheme and whether these purposes have helped the advancement of the FCA test case. The purpose of the Scheme, in sum, is to provide immediately relevant authoritative guidance in areas of general importance to English law and — with regard to the FCA test case — for any such guidance to clarify matters of contractual uncertainty. This chapter has confirmed that the Scheme was a seemingly excellent choice of means by which to resolve the disputes. However, the use of the Scheme in the FCA test case can only be deemed successful if all the disputes were met with satisfactory outcomes, per the objectives of the Scheme in the FCA test case, which will be explored in the next chapter.

CHAPTER II – THE DISPUTES

As explored in the preceding section, the Scheme is *prima facie* an excellent choice through which the FCA test case could be resolved. It utilises an expedited process for decisions, one of the main reasons the Scheme was proposed and utilised to begin with. This section seeks to ascertain whether this

suitability has translated into satisfactorily-resolved disputes. Additionally, by examining the actual issues and disputes that were dealt with by the Court, this article explores whether the usually uncompromising quality of Supreme Court judgments has been lost to make way for this hastened process of litigation.

The ensuing discussion does not attempt to comprehensively assess or evaluate the substantive decisions surrounding the various areas of dispute. As such, the substantive law will not be considered in any great detail. It merely assesses whether each dispute in question was satisfactorily resolved. Firstly, it will be considered whether immediately relevant authoritative English law guidance was provided, per the objective of the Scheme. Following this, it will be considered whether any such guidance has clarified the key issues of contractual uncertainty, as per the objective of the FCA test case overall. The first dispute considered is related to disease clauses, followed by the prevention of access/hybrid clauses, causation, trends clauses, and an evaluation of the *Orient Express Hotels*³⁹ case. It should be noted that the use of the word ‘dispute’ is not entirely accurate: the Scheme was used to provide guidance on uncertain areas of law. Thus, the following discussion will not relate to actual disputes as such, but the word ‘dispute’ will be used to describe the relevant areas of uncertainty.

(1) THE DISEASE CLAUSES

‘Disease’ clauses are used to provide non-damage business interruption cover, which, in the case of the sample policy wordings used, are triggered by an occurrence of a notifiable disease within the vicinity or a specified radius of the insured premises.⁴⁰ The determination by the Supreme Court honed in on a standard disease extension clause, one which provides cover for ‘any ... occurrence of a Notifiable Disease within a radius of 25 miles of the Premises’.⁴¹ For these purposes, ‘Notifiable Disease’ was defined by the policy as ‘illness sustained by any person resulting from ... any human infectious or human contagious disease ... an outbreak of which the competent local authority has

³⁹ *Orient-Express Hotels* (n 9).

⁴⁰ ‘Covid-19: Business Interruption Insurance’ (*The Law Gazette*, 2021) <<https://www.lawgazette.co.uk/law-reports/covid-19-business-interruption-insurance/5107125.article>> accessed 2 August 2021.

⁴¹ *Financial Conduct Authority* UKSC (n 7) [50], [54], [61], [88], [93].

stipulated shall be notified to them'.⁴² The Supreme Court agreed with the insurers' interpretation of the scope of cover provided under 'disease' clauses of this nature, namely that they cover only relevant effects of COVID-19 cases that occur at, or within, the specified radius of a premises.⁴³ However, this determination by the Supreme Court proves somewhat futile, as will be discussed.

The Supreme Court rejected the High Court's decision that the insured peril was, in effect, COVID-19. The Supreme Court's analysis focused on the ordinary, plain English language meaning of the words used.⁴⁴ In particular, the meaning of 'occurrence of a Notifiable Disease' referred to the well-established meaning of the word 'occurrence' in insurance law (that is, something that happens at a particular time and place and in a particular way).⁴⁵ In reaching this view, the Supreme Court considered that the High Court had erred in its exercise of interpretation and had wrongly taken into account two 'fundamental' issues, which the Justices considered were relevant solely to the separate issue of causation.⁴⁶ The first issue concerned the High Court's observations that a typical 'disease' clause does not expressly confine cover to business interruption resulting only from cases of a notifiable disease within the specified area, as opposed to cases elsewhere. The second issue was that the nature of a notifiable disease meant that it had the potential to affect a wide area and prompt the authorities to take action in response to an outbreak as a whole, as opposed to localised parts of it.⁴⁷ In this case, there was no such ambiguity, as no reasonable reader would understand the words 'any occurrence of a Notifiable Disease within a radius of 25 miles' to include the occurrence of such disease outside that radius. To decide otherwise would be to 'stand the clause on its head', and so, the plain English meaning of the words was given effect.⁴⁸ This interpretation resultantly ensured that many policyholders were able to recover compensation, owing to the fact that COVID-19 was a disease that reached the entire country.

⁴² *ibid* [51].

⁴³ *ibid* [210].

⁴⁴ *ibid* [67].

⁴⁵ *Axa Reinsurance (UK) Ltd v Field* [1996] 1 WLR 1026 (HL); *Kuwait Airways Corp v Kuwait Insurance Co SAK* [1999] UKHL 12; *Mann v Lexington Insurance Co* [2001] 1 Lloyd's Rep 1 (CA).

⁴⁶ *Financial Conduct Authority* UKSC (n 7) [56].

⁴⁷ *ibid* [52]-[56].

⁴⁸ *ibid* [61].

In any event, the Supreme Court considered the overriding question to be how the words of the contract would be understood by a reasonable person.⁴⁹ In the case of an insurance policy sold principally to SMEs, the person to whom the document should be taken to be addressed is not a pedantic lawyer who will subject the entire policy wording to minute textual analysis. It is an ordinary policyholder who, on entering into the contract, is taken to have read through the policy conscientiously in order to understand what cover they were getting.⁵⁰

These considerations by the Supreme Court would suggest that the Scheme was successful for this aspect of the ruling. The overriding objective of the Scheme was to provide immediately relevant authoritative English law guidance. The Supreme Court provided such when it overruled the decision of the High Court. The decisions of the High Court are innately less authoritative than Supreme Court decisions, and so the quick appeal allowed for a more authoritative stance to be provided.

The Supreme Court expressly stated that the guidance aimed to be clear and concise for the benefit of the ordinary policyholder and was to be provided in an unusually short period of time.⁵¹ It eliminates any controversy with regard to disease clauses of this nature. Moreover, this ruling provides guidance which clarifies a key issue of contractual uncertainty. Thus, the Scheme has successfully operated to largely resolve this issue of disease clauses. Whilst this will greatly assist many aggrieved policyholders, there is scope for future litigation if policyholders and insurers construe the Supreme Court's guidance differently. There is no doubt that immediately relevant authoritative English law guidance was provided, but it is questionable to what extent this guidance clarified the issues of contractual uncertainty in relation to the disputes around disease clauses, as further disputes could arise from differing interpretations of the ruling.

⁴⁹ Peter Blakemore, 'Back to Basics: Drafting Pandemic Clauses' [2021] 2114 Estates Gazette 48-49.

⁵⁰ *Financial Conduct Authority* UKSC (n 7) [77].

⁵¹ *ibid.*

(II) PREVENTION OF ACCESS CLAUSES AND HYBRID CLAUSES

The aptly named ‘hybrid clauses’ are an amalgamation of both disease clauses, as discussed above, and prevention of access clauses, which are clauses that provide cover where some form of authority so acts or there is some occurrence that prevents or restricts access to, or use of, the insured premises. At first instance in the High Court, the FCA was semi-successful in arguing that such clauses would provide cover for the pandemic and its resulting consequences.⁵² On appeal, the Supreme Court was tasked with resolving three issues. The first of these was whether the public authority measures were imposed by law. The Supreme Court affirmed the ruling of the High Court in stating that ‘restrictions imposed’ was tantamount to a legally binding compulsive instruction.⁵³ Conversely, though, the Supreme Court rejected the notion that all restrictions refer to the use of legal powers.⁵⁴ The test for interpreting the relevant words was how they would be understood by the standards of a reasonable person.

The Supreme Court did not consider, however, the individual circumstances of each measure, both general and specific, with regard to this reasonable person test. Furthermore, there was no consideration of whether the measures were set out in a reasonable way that allowed for an easy determination of precisely what satisfactory compliance required.⁵⁵ The Supreme Court should have gone further by expressly stating how the policyholders could comply with their policies, and so it is submitted that the process of interpretation was not thorough enough. This represents a failure of the use of the Scheme. Naturally, expediting any process will require truncation. Yet, it would appear that by omitting any discussion about what a reasonable person may understand by each measure, there is scope for further confusion.

There could be future claims arising from thwarted interpretations of these restrictions, something which could have been avoided through a more

⁵² John Wright, ‘High Court Test Case - Not Conclusive’ [2020] 31 (10) Construction Law 30-32.

⁵³ *Financial Conduct Authority* UKSC (n 7) [112]-[116].

⁵⁴ *ibid* [120].

⁵⁵ Stuart Pemble, ‘Does Business Interruption Insurance Cover Covid-19?’ (2020) 2043 *Estates Gazette* 45.

thorough consideration of the individual measures.⁵⁶ Whilst this was not feasible within the FCA test case timeline, per the usage of the Scheme, normal litigation would have likely provided ample time to do so. Litigation would have facilitated not only a far more thorough examination of how the policies would be understood but could have also facilitated contributions from policyholders to ensure that ordinary people interpreted the policies correctly. This would have decreased the likelihood of future litigation. There may have been immediately relevant authoritative English law guidance provided, but this guidance did not clarify the key issues of contractual uncertainty surrounding the ‘imposed by law’ requirement.

The second of these issues concerns the level of interruption to the business for cover to be triggered. The analysis primarily centred around extensions relating to the ‘inability to use’, ‘prevention of access’ and ‘interruption’. With ‘inability to use’, the Supreme Court did not depart from the ruling of the High Court that a mere hindrance or impairment of use would be satisfactory, further qualifying the ruling by expressing that there was no requirement for interruption to the entirety of the premises for all purposes.⁵⁷ This requirement would therefore be satisfied even where only a ‘discrete’ part of a business’ premises or activities were interrupted.⁵⁸ This is a good example of both the success and the failures of the FCA test case. This facet of the judgment has not provided an express answer to the issues surrounding ‘inability to use’ clauses. However, it has provided the means for doing so. As a fact-dependent inquiry would be necessary on a case-by-case basis to determine whether the tests have been satisfied, it would take a far longer period of time to establish whether, for example, a business has not been able to operate a ‘discrete part of its business’. However, the ruling provided immediately relevant authoritative English law guidance and offered some degree of clarity within an especially short period of time.

The Supreme Court further held that the wording of ‘prevention of access’ clauses was to be construed to mean that there was a wholesale prevention of the use of the premises, as distinct from mere hindrance.⁵⁹ This

⁵⁶ *Financial Conduct Authority* UKSC (n 7) [61].

⁵⁷ *ibid* [128]-[145].

⁵⁸ *ibid* [127], [281].

⁵⁹ *ibid* [146]-[152].

falls neatly in line with the above analysis regarding ‘inability to use’. The ‘interruption’ extensions were interpreted as including disruption, however slight, which may not entirely close the business and its operations, so long as there is a material effect on the finances of the business in question.⁶⁰ Moreover, for the third issue, it was for the Supreme Court to determine whether the restrictions imposed needed to be directed at the policyholder and not the general public, per the insurers’ submissions. The Supreme Court rejected this notion⁶¹ and subsequently found that there was no requirement for the restrictions to be directed at the policyholder.⁶² This determination was made in adherence with the aims of the FCA test case. This part of the ruling is clear and provides immediately authoritative English law guidance. Moreover, the guidance provides clarity to an issue of contractual uncertainty, and is satisfactory in that policyholders will be able to claim for cover in situations where, prior to the ruling, cover was not granted.

Overall, it can be asserted that the determinations of the Supreme Court operate overwhelmingly in favour of policyholders and have overruled the High Court on a number of issues. It has interpreted prevention of access and hybrid clauses considerably more widely than the High Court at first instance. These rulings are also made in line with the objectives of both the FCA test case and the Scheme. Immediately authoritative English law guidance was provided through declaratory relief in a friendly action in a manner that clarified key issues of contractual uncertainty. Thus, it can be established that this part of the judgment was mostly successful.

(III) CAUSATION

Due to the causative nature of insurance contracts, causation is often the trickiest issue that arises during insurance litigation, and this trend continued in the FCA test case. The most pertinent issue before the Supreme Court centred around the interpretation and resultant application of the infamous ‘but-for’ test.⁶³ The insurers’ submissions advanced the proposition that the

⁶⁰ *ibid* [158].

⁶¹ Ed Hundleby, ‘FCA Business Interruption Test Case - Lessons’ (2021) *Pensions Today* 4-5.

⁶² *Financial Conduct Authority* UKSC (n 7) [128].

⁶³ Mexian Song, ‘Revisiting Concurrent Causation And Principles In English Insurance Law: A Legal Fiction’ (2021) 6 *Journal of Business Law* 457-474.

policyholder needed to demonstrate that the loss would not have been suffered but for the occurrence of the insured peril.⁶⁴ The Supreme Court sternly rejected this notion. In doing so, they inadvertently affirmed the usefulness of the Scheme. The Justices undertook a whirlwind revision of not only proximate causation but also of the law of concurrent causation in an extraordinarily short time period.⁶⁵ They expressly acknowledged how, since the 19th century, proximate cause has not been concerned with the cause most proximate from the loss in terms of time, but refers instead to the efficient cause of the loss.⁶⁶ This test was devised for the benefit of the reasonable, average man, not for a scientist or any other person who would consider different interpretations of the word proximate.⁶⁷

The Supreme Court did note, however, that this is not a matter of ‘unguided gut feeling’; rather, it is usually determined through various principles that have emerged from centuries of jurisprudence.⁶⁸ The Court would typically consider whether the peril covered had any causal involvement in the loss; whether any excluded perils had a part in the loss; and whether in the event of concurrent losses, there is covered loss or not.⁶⁹ In these more common types of cases, the peril covered was still a cause in the causal chain of events. The differentiator in the FCA test case is that the Supreme Court considers each case of COVID-19 to be a separate occurrence and insured peril.⁷⁰ The Supreme Court expressly acknowledged that, as per their own interpretations, there were not to be merely two or three concurrent causes of the loss suffered, but a vast number, possibly stretching into the millions, depending on the precise number of known infections in the country at the time restrictions were introduced by the UK Government.⁷¹

⁶⁴ *Financial Conduct Authority* UKSC (n 7) [238]; *ibid.*

⁶⁵ *ibid* [162]–[171].

⁶⁶ *ibid* [165]–[167]; See *Reischer v Borwick* [1894] 2 QB 548, 550; *Leyland Shipping Co Ltd v Norwich Union Fire Insurance Society Ltd* [1918] AC 350, 365 (Lord Atkinson); *Becker, Gray & Co v London Assurance Corpn* [1918] AC 101, 113–114.

⁶⁷ *Yorkshire Dale Steamship Co Ltd v Minister of WarTransport* [1942] AC 691 per Lord Wright at 706; Joseph Arnould and others, *Arnould On The Law Of Marine Insurance And Average* (20th edn, Sweet & Maxwell 2021).

⁶⁸ *Financial Conduct Authority* UKSC (n 7) [168].

⁶⁹ *ibid.*

⁷⁰ *ibid.*

⁷¹ *ibid* [201].

Once this line of reasoning had been established, it was for the Supreme Court to determine whether the causal requirements for the different clauses were satisfied. With regard to the disease clauses, the Supreme Court concluded that it could not be the commercial intentions of the parties to treat any uninsured cases of COVID-19 occurring outside the radius of the policy to preclude a policyholder of indemnity in relation to business interruption cover.⁷² The Supreme Court saw this interpretation as not only relatively clear and simple, but also as constituted in accordance with the overarching need for clarity and certainty.⁷³ It is submitted that the Supreme Court's analysis here is correct: they have adhered to the objectives of both the Scheme and the FCA test case as a whole. A slightly different interpretation was adopted for the prevention of access and hybrid clauses as compared to these clauses, where more than one condition needs to be satisfied to conclude that an interruption to the business has been caused by an insured peril.⁷⁴ The Supreme Court held that there were two or more causes of the loss — the covered peril as described in the policy as well as the pandemic and broader social considerations — each of which would have caused the loss. As per the correct interpretation of the policy, the requisite causal connection could be satisfied by finding that each case of COVID-19 was 'a separate and equally effective cause of the restriction'.⁷⁵

(IV) TRENDS CLAUSES

Trends clauses are provisions that dictate the relative adjustment of an insured loss to ensure, *inter alia*, that the indemnity granted is not unjustifiably increased for uninsured reasons. As noted in the judgment, 'whilst the basic comparison between the turnover of the business in the prior period and in the indemnity period will produce a rough quantification of the lost revenue, there may be specific reasons why a higher or lower figure would be expected for the indemnity period apart from the operation of the insured peril'.⁷⁶

⁷² *ibid* [237].

⁷³ Joanna McCunn, 'Fallen Arch: Construction, Causation And Covid-19' (2021) 37 (2) *Professional Negligence* 85-89.

⁷⁴ *ibid*.

⁷⁵ *ibid*.

⁷⁶ *Financial Conduct Authority UKSC* (n 7) [254].

The purpose of trends clauses is to provide for adjustments to be made to reflect ‘trends’ or ‘circumstances’ that result in an inflated or deflated figure.⁷⁷ In the FCA test case, the Supreme Court had to establish the extent to which, if at all, the insurers could reduce the indemnity as, if the policyholder’s business had managed to remain open it would have suffered a downturn in revenue due to the effects of the COVID-19 pandemic.⁷⁸

At paragraph 262, the Supreme Court expressly denotes its objective of striving to identify an interpretation of the trends clauses aligned with the insuring clauses ‘so as not to take away the cover provided by the insuring clauses’.⁷⁹ This consideration operates only to advantage policyholders. It was the opinion of the Supreme Court that the most unproblematic means by which the trends clauses should be construed is, in the absence of any provisions to the contrary, by acknowledging that the objective of such clauses is to consider the result that would have occurred but for the insured peril and the circumstances stemming from the same originating cause.⁸⁰ Conversely, the insurers’ submissions would see a quantification procedure transform into an exclusionary procedure.⁸¹ Subsequently, there must be a complete omission of the numerous COVID-19 consequences when calculating the success of the business had the insured peril not occurred. This determination is consistent with the earlier discussion of causation.⁸² It further provides a clear route through which policyholders can claim indemnity where it was previously unavailable or substantially reduced by insurers.⁸³ Here, clear authoritative English law guidance has been provided, clarifying key issues of contractual uncertainty concerning trends clauses. It is an important aspect of the overall judgment and plays into the objectives of the FCA test case.

⁷⁷ *ibid* [253]-[255], [284].

⁷⁸ *ibid* [289]-[294].

⁷⁹ *ibid* [262].

⁸⁰ *ibid* [268].

⁸¹ *ibid* [146]-[152].

⁸² *ibid* [168].

⁸³ Harris J, ‘Supreme Court Judgment Favours FCA In Business Interruption Insurance Appeal’ (2021) 42 (4) *Company Lawyer* 122.

(V) ORIENT EXPRESS HOTELS

The Supreme Court's treatment of the watershed High Court case, *Orient Express Hotels*,⁸⁴ was highly anticipated for its potential resounding effect on the future insurance market. In *Orient Express Hotels*, a claim was advanced for business interruption to a hotel arising from a hurricane under an all-risks policy. This claim was subsequently rejected as the 'but-for' test was interpreted as meaning that the business interruption damage would have occurred regardless due to the fact the area as a whole had been subjected to substantial hurricane damage.⁸⁵ The interpretation considered what precisely would have happened if the damage had not occurred, but the hurricanes had. The insurers used this judgment to submit that, in the contemporary context of the COVID-19 pandemic, the policyholders would have suffered a loss notwithstanding the Government impositions, as the pandemic would have caused a loss in revenue.⁸⁶

The Supreme Court rejected this analysis and expressly overruled *Orient Express Hotels*.⁸⁷ It held that, in a scenario where both the insured peril, as well as the uninsured peril, operate concurrently, and from this concurrent operation the same fortuity is produced, then so long as the damage which is caused by the operation of the uninsured peril is not expressly or otherwise excluded from the policy, the loss stemming from both of these causes is covered.⁸⁸ This outcome will certainly see enormous implications for insurers and policyholders alike for the adjustment of property damage, as well as for non-damage business interruption claims.⁸⁹ This ruling massively relaxed the law established in *Orient Express Hotels*.⁹⁰ It was limited in scope as it was made under the aegis of the Arbitration Act 1996, s.69, which is limited to questions 'arising

⁸⁴ *Orient-Express Hotels* (n 9)

⁸⁵ N.B. the case was heard by Hamblen J (as he then was) on appeal from a decision of an arbitral tribunal (which included Mr George Leggatt QC, as he then was). Ten years later, Lord Hamblen and Lord Leggatt are now Justices of the Supreme Court and gave the main judgment in the FCA test case.

⁸⁶ 'Test Case Reassures On COVID Insurance' (*New Law Journal*, 2021)

<<https://www.newlawjournal.co.uk/content/test-case-reassures-on-covid-insurance>> accessed 11 March 2023.

⁸⁷ *Financial Conduct Authority* UKSC (n 7).

⁸⁸ *Orient-Express Hotels* (n 9) [306]-[308].

⁸⁹ *Financial Conduct Authority* UKSC (n 7).

⁹⁰ *ibid.*

out of an award’.⁹¹ Additionally, as it was an arbitration award, there was far less scope for a detailed argument; in contrast, despite its expedited process, the FCA test case allowed for a far more thorough consideration of the matter. Moreover, this ruling adheres to both the objectives of the FCA test case as well as the Scheme. Immediately authoritative English law guidance has been provided, with this guidance clarifying the contractual uncertainty stemming from the *Orient Express Hotels*⁹² decision. The resulting position is clearer, and thus far more satisfactory, than before.⁹³

In summary, it can be generally stated that the preceding disputes were resolved satisfactorily, with some minor exceptions. In all five of the main disputes, immediately relevant authoritative English law guidance was provided. However, in some determinations, this guidance does not necessarily provide clarity on some contractual uncertainty. Guidance provided for the disease clauses clarified issues of contractual uncertainty, namely, how to construe such clauses. With prevention of access and hybrid clauses, three matters required resolution. The first issue on ‘imposed by law’ was met with authoritative guidance, but this guidance fell short of wholly clarifying matters of contractual uncertainty. A longer process, perhaps through litigation, would likely have resolved this, highlighting a disadvantage of using the Scheme. This judgment has provided the basis for a negotiated settlement and the FCA test case has subsequently provided updates on settlement in this area. However, the issue was not competently dealt with by the FCA test case, most likely due to the expedited process of the Scheme. The other two issues – namely, interruption extensions and whether impositions need to be ‘directed at the policyholder’ – were resolved soundly, with the guidance provided clarifying issues of contractual uncertainty.

The causation issues were also satisfactorily answered. The Supreme Court offered immediately relevant authoritative English law guidance. This guidance resolved the issues of contractual uncertainty, especially those relating to the interpretation of the ‘but-for’ test and the proximate cause of the loss. There was a similar level of success with the trends clauses, with the guidance clarifying areas of contractual uncertainty by declaring that the effects of

⁹¹ Arbitration Act 1996, s 69(1).

⁹² *ibid.*

⁹³ *Financial Conduct Authority UKSC* (n 7) [253]-[255], [284].

COVID-19 were not to be considered when quantifying cover. The final issue, that of *Orient Express Hotels*,⁹⁴ was also resolved successfully. The guidance overruled the judgment, clarifying the position for many insureds and ensuring many more policyholders will be entitled to cover.

Overall, this chapter has demonstrated that the disputes were soundly and successfully resolved per the objectives of the Scheme in the FCA test case. In the vast majority of the rulings, there was immediately relevant authoritative English law guidance provided. Moreover, this guidance has largely clarified key issues of contractual uncertainty. This would denote that the use of the Scheme in the FCA test case, per the objectives of the Scheme, was a success. However, the judgment is lacking in some areas, and a more thorough consideration of some of the issues could have resulted in an even more satisfactory outcome. For example, a minute analysis of the extent to which damages would be payable would have improved the judgment. If there had been more robust and complete guidance, it would have resulted in more of the issues of contractual uncertainty being resolved, but this was not possible in the truncated timeframe. Thus, perhaps the Scheme was not the most optimal method by which the issues ingrained in the FCA test case could have been resolved. This will be explored in the next chapter.

CHAPTER III – THE SUCCESS

(1) EVALUATING THE SUCCESS OF THE SCHEME IN THE FCA TEST CASE

The first part of this section will provide an overall assessment of the degree of the Scheme's success. It will do so by tackling several research questions. It will explore whether the Scheme achieved its purpose, as well as whether the objectives behind initiating the FCA test case were met. Furthermore, it will assess whether all of the issues were satisfactorily resolved, and further whether there were any new issues created by the judgment. The use of Scheme in the FCA test case can only be judged as successful if the objectives set out at the preliminary stages of the action have been met.⁹⁵ As has been continuously noted throughout this article, the objective of the use of the

⁹⁴ *ibid.*

⁹⁵ *ibid.*

Scheme in the FCA test case was to provide immediately relevant authoritative English law guidance, with any such guidance needing to clarify key issues of contractual uncertainty.⁹⁶ To this end, the Scheme was largely successful and the use of it in the FCA test case is an excellent endorsement for the continuation of the Scheme, despite being used for the first time. It is an extraordinary feat for a claim of such magnitude to progress from commencement to the Supreme Court in a little over seven months. The Supreme Court noted that ‘this determination will facilitate prompt settlement of many claims and achieve considerable savings in time and cost’.⁹⁷ However, it is this ‘considerable savings in time and cost’ which has the potential to give rise to future issues.

The individual disputes were all dealt with in a satisfactory standard, per the objective of the use of the Scheme in the FCA test case. However, whilst clarificatory guidance was provided in most of the relevant disputes, for some the determinations gave rise to future uncertainty.⁹⁸ With disease clauses, the Supreme Court’s ruling described how to interpret relevant policy wordings. Whilst this will greatly assist many aggrieved policyholders, there is larger scope for potential future litigation than for other clauses if policyholders and insurers construe the Supreme Court’s guidance differently. The Supreme Court’s rulings in relation to both prevention of access and hybrid extensions (such as the determination that some measures which were not imposed under the force of law could still trigger coverage) could cause additional claims to be brought or increase indemnities being paid on individual claims. The judgment of the Supreme Court will not be applicable to all scenarios and thus, these excluded scenarios will need to be identified and adjusted on a case-by-case basis. Whilst individual litigation would have taken significantly longer and been significantly more expensive, it would have resulted in a finalised interpretation of the relevant policy without any issues.

The use of the Scheme in the FCA test case also illuminated other disadvantages compared to litigation. At the preliminary stages, there were 16 insurers invited to participate in the test case. Out of these 16, only 8 accepted

⁹⁶ *ibid.*

⁹⁷ *ibid* [43].

⁹⁸ Becca Hogan, ‘Partner Becca Hogan examines the under-utilised collective redress tool’ (*Signature*, 21 June 2001)

<<https://www.signaturelitigation.com/partner-becca-hogan-examines-the-under-utilised-collective-redress-tool/>> accessed 1 September 2021.

the invitation, largely due to the probable unfavourable outcome for insurers.⁹⁹ This rejection of admittance onto the FCA test case represents a negative aspect of the Scheme. Parties join through a process of mutual agreement and advance on the basis of an agreed set of facts, or in this case, policy wordings. If a party does not agree to be involved in the test case, there can be no guidance provided by the court in relation to that party. This resulted in the High Court (as well as the Supreme Court once appealed) working with sample policy wordings which were statistically only representative of half of the insurers in the business interruption insurance market. Thus, the judgment given cannot necessarily be utilised as a universal reference point for determining the outcome of claims for business interruption cover. This exemplifies a situation where litigation would be advantageous instead as actors cannot disagree to be a party to a litigation. The issue with litigation, however, is that there is a need for a real and present dispute between the parties for an action to proceed. When using the Scheme, a party can commence a pre-emptive action prior to an actual dispute arising — one of the Scheme's most attractive features. If it was necessary for a dispute to arise before parties could commence an action, more significant loss and harm would have to occur before guidance can be provided.

Moreover, the FCA was arguably over-selective with the choices of policy wordings they put before the High Court. All of the policy wordings considered were from larger insurers in the market. As mentioned in the preceding paragraph, several very large insurers were not invited or had their invitations rejected, such as AXA and Aviva. There was also an extremely large number of smaller insurers who were excluded altogether. The FCA approximates that there are around 700 different types of business interruption policies in use in the UK insurance market from around 60 different insurers.¹⁰⁰ Given that the judgment considers only specific policy wordings, there are likely to be further disputes between insurers and policyholders that may result in litigation. Situations like these render the use of the Scheme somewhat futile.

Whilst it is accepted that litigation involving all these insurers and policyholders is unreasonable in terms of time and cost, it is submitted that the

⁹⁹ 'Update on FCA Test Case of the Validity of Business Interruption Claims' (FCA, 1 June 2020)

<<https://www.fca.org.uk/news/press-releases/update-fca-test-case-validity-business-interruption-claims>> accessed 10 March 2023.

¹⁰⁰ *ibid.*

FCA test case could have considered a wider range of sample policy wordings to ensure the guidance provided in the judgment would apply more universally. It is acknowledged that this is asking a lot of an expedited mechanism, but whilst dealing with such important issues, this request can be justified. The Scheme's objective was to provide 'immediate' guidance. If there is litigation commencing after the Supreme Court judgment on the correct interpretation of a policy, it would have been quicker just to litigate in the first place. Nevertheless, insurers will now be aware that they have far less room for manoeuvre with regard to interpretation than before the FCA test case.

Overall, the judgment is beneficial to policyholders and undoubtedly puts them in a better position than before the FCA test case. It is submitted, however, that it is questionable just how much better this position is. The use of the Scheme has allowed for immediately relevant authoritative English law guidance to be provided for in all of the disputes considered. As explored above, this guidance was provided in an extraordinarily short period of time, albeit perhaps not 'immediately' in light of the appeal to the Supreme Court. Moreover, the guidance was relevant to the disputes and authoritative in its nature. In this respect, the use of the Scheme was successful. It achieved its overarching objective whilst simultaneously saving enormously on time and, consequently, on cost as well. The concept of running a case without a 'real dispute' has proven to work incredibly well. It has allowed for guidance to be provided in relation to a dispute which would likely arise in the future, and thus could represent the beginning of a paradigm shift in the way litigation is conducted in England and Wales. This will be explored in the next section of this chapter.

However, whilst the primary objective of the Scheme was largely achieved — namely, to provide immediately relevant authoritative English law guidance — the primary objective of the test case as a whole was not quite as closely met. The FCA test case strived to clarify key issues of contractual uncertainty. This was achieved only to a limited degree. Some areas of dispute, such as causation, were indeed largely clarified. However, in other areas of dispute (such as disease clauses, prevention of access clauses and hybrid clauses) the attempted clarification inadvertently gave rise to potential future uncertainty. Whilst the guidance provided will greatly assist many aggrieved policyholders, there is scope for potential future litigation if policyholders and insurers

construe the Supreme Court's guidance differently. As has been previously mentioned, any litigation stemming from the guidance would render the use of the Scheme in the FCA test case wholly futile. Any resultant litigation would take a significant period of time, especially if there are appeals. In addition, if the guidance provided does not clarify areas of uncertainty but rather causes further uncertainty resulting in litigation, then the use of the Scheme in the FCA test case cannot be judged to be entirely successful.

(II) TEST CASES – THE FUTURE OF CIVIL DISPUTE RESOLUTION?

This final section will look at the wider context of the future use of the Scheme. It will explore whether the Scheme has improved financial services dispute resolution or whether the pre-existing methods, such as ordinary litigation, are better. It will look at whether the Scheme should be extended to other areas of law or whether it should be restricted to financial services. This section will propose improvements to the Scheme in addition to assessing its disadvantages. Consideration of the above will assist in making a contemporaneous evaluation of the degree of success of the Scheme in the seminal FCA test case.

It is clear, however, that the Scheme is here to stay, at least for the resolution of similar disputes. It was promoted to a permanent fixture of Practice Direction CPR Part 63A on the 1st of October 2020.¹⁰¹ Yet, despite the fairly successful use of the Scheme in the FCA test case, it must be applied cautiously in the future. As previously discussed, the FCA test case affected many insurers and policyholders, none of whom were actually party to the proceedings. The whole test case was heard in a vacuum devoid of human emotion. There was a complete absence of any opportunity for either the High Court or the Supreme Court to hear from aggrieved policyholders whose businesses were financially devastated by the COVID-19 pandemic, or by employees whose livelihoods have been destroyed.¹⁰² This was to ensure that the whole process could be concluded as quickly as possible, in order for the Court to then provide immediate guidance. It is submitted that whilst the judgment was largely, but not wholly, in favour of policyholders, the guidance could have been different had these people been heard. This would have made the

¹⁰¹ Hogan (n 98).

¹⁰² *Financial Conduct Authority* UKSC (n 7).

judgment more relatable and accessible to ordinary people, as they could compare the fact-pattern heard in the case to their own situations. It is accepted that aspects of normal litigation will have to be disposed of in this expedited process for the sake of speed, but the inclusion of an opportunity for laypeople to be heard would be beneficial to the judgment overall despite the consequent delay.

The guidance provided in the judgment has also been written with little regard to laypeople affected by it, given that it is written in such a jargon-filled sesquipedalian manner that it is an impossible task for laypeople to understand it.¹⁰³ While true that the case involves incredibly complicated issues and the judgment was always going to require the communication of nuanced ideas, it is submitted that there could have been more consideration made by the court to the people who will be affected by the judgment. It is important for the future of the test case procedure to consider the needs of a person's access to justice by allowing individual voices to be heard where appropriate.¹⁰⁴ This will allow for more tailored guidance to be provided, which will benefit the ordinary people for whom the Scheme was invented.

Similar sentiments have been expressed in *Rolls Royce PLC v Unite the Union*,¹⁰⁵ a case which is viewed by some as a sort of 'precursor' to the FCA test case. In this case, the parties disputed whether the inclusion of length of service within a selection matrix for redundancy purposes would amount to unlawful age discrimination. The court was asked whether it was correct to make a declaratory judgment when the case had otherwise been effectively settled. In doing so, the court provided guidance as to the interpretation of The Equal Treatment Directive in an omission of any actual dispute between the relevant parties. Relevant to this discussion in particular was the recognition that the guidance would affect a significant number of people, yet none of these people were heard in the case.¹⁰⁶ The judge at first instance, Sir Thomas Morison, recognised this flaw and expressed his disdain at the situation, stating that he

¹⁰³ Lucy Pert, 'The FCA Test Case And Collective Justice - The Jury Is Out' (Hausfeld, 23 February 2021)

<<https://www.hausfeld.com/en-gb/what-we-think/perspectives-blogs/the-fca-test-case-and-collective-justice-the-jury-is-out/>> accessed 5 September 2021.

¹⁰⁴ *ibid.*

¹⁰⁵ *Rolls-Royce* (n 37).

¹⁰⁶ *ibid*; Hogan (n 98).

had ‘considerable misgivings’ about the proceedings omitting ‘the advice and wisdom’ of lay members or any oral evidence from lay members at all.¹⁰⁷

The court made it clear that they endeavoured to avoid any future confusion or scope for dispute by interpreting the direction in an accessible way. No similar considerations were made in the FCA test case. The *Rolls Royce PLC v Unite the Union* case additionally demonstrates that some of the aspects of the Scheme are not wholly original but have been seen in previous cases. *Rolls Royce PLC v Unite the Union* was heard at least five years prior to the consideration of the setting up of the Financial List in which the Scheme resides, yet it exemplifies a case being heard without an actual dispute having arisen and without the direct involvement of any aggrieved parties.¹⁰⁸ It can thus be stated that the Scheme encompasses aspects of civil procedure which were already in use and has merged them into a more defined procedure. This neatly codifies the Scheme and is therefore a welcome advantage to its creation and use.

This Scheme and the procedures included therein are not the only means of financial services dispute resolution outside the traditional route of litigation. The Financial Ombudsman Service (hereinafter FOS) resolves thousands of financially-centred disputes annually whilst retaining the advantages of the scheme.¹⁰⁹ The FOS does this whilst still retaining many of the advantages of the Scheme. It does not incur nearly as much cost as litigation, does not involve the innate procedural complexity of litigation, and takes a considerably shorter period of time.¹¹⁰ Yet, it is not perfect and has several flaws that are not found within the Scheme. There is no scope or platform for the FOS to engage with external stakeholders in a case, such as the FCA, when passing a judgment.¹¹¹ Thus, the FOS is very well-suited to resolving fact-specific, individual complaints, but much less suited to dealing with cases that involve a large number of disputes, the resolution of which have wide implications. Additionally, there is no appeals process ingrained into the FOS, making it even less suited to reconcile larger-scale financial disputes as there is

¹⁰⁷ *Rolls-Royce plc* EWCA (n 37) [7].

¹⁰⁸ *ibid*; Hogan (n 98).

¹⁰⁹ ‘Complaints We Can Help With’ (*Financial Ombudsman*, 2021)

<<https://www.financial-ombudsman.org.uk/consumers/complaints-can-help>> accessed 14 September 2021.

¹¹⁰ *Orient-Express Hotels Ltd* (n 9) [3].

¹¹¹ *ibid*.

less scope for scrutiny of far-reaching decisions.¹¹² The Scheme does not suffer from any of these flaws in the FOS, or those ingrained into litigation. It is not even remotely comparable to litigation in terms of cost or speed, with the Scheme being far cheaper and faster even with an appeal. Thus, it is submitted that, to a fairly substantial degree, the inception of the Scheme has improved financial dispute resolution in England.

The relative success of the Scheme has been recognised by practitioners across the legal profession, and it has been suggested that it could be extended to other types of disputes with significant importance to the market.¹¹³ For example, the Scheme could be utilised to resolve similar disputes arising from the aftermath of the COVID-19 pandemic or even to resolve an avalanche of issues inevitably arising from Brexit.¹¹⁴ It should be noted that as the FCA test case was the first to be heard under the Scheme, it cannot be concluded that the Scheme's implementation has been overall positive. For that to be concluded, there would need to be other cases which utilise the Scheme at least to the same level of success as in the FCA test case. The FCA test case was a good endorsement for the continued use of the Scheme, but as has been discussed earlier in this article, it was far from perfect.

It can be stated that the utilisation of the Scheme in the FCA test case was largely successful. The overarching objective of the Scheme — namely, to provide immediate relevant authoritative English law guidance — was met in all of the disputes at hand. Moreover, the overarching objective of the FCA test case — namely, to assist in clarifying key issues of contractual uncertainty — was also met, but to limited degrees in some of the disputes. The FCA test case has proven to primarily be a good endorsement, and an example of the advantages of the Scheme when compared to litigation or the FOS, but it has also demonstrated disadvantages which must be considered. Overall, it can be said that the Scheme has improved financial services dispute resolution, given its advantages over the alternative pre-existing methods.

¹¹² *ibid.*

¹¹³ Pert (n 103); Blundell, Donnelly and Emson, 'Financial Services And Markets Dispute Resolution Quarterly Update: October 2020' (*Macfarlanes.com*, 2021) <<https://www.macfarlanes.com/what-we-think/in-depth/2020/financial-services-and-markets-dispute-resolution-quarterly-update-october-2020/#Financial>> accessed 3 September 2021.

¹¹⁴ Financial Services (n 113).

CONCLUSION

The COVID-19 pandemic has given rise to unthinkable social, medical, and economic ramifications. Fortunately, seldom does a situation arise where the livelihoods of hundreds of thousands of people are contingent upon the judgment of a single case. There was an urgent need for resolutions to assist people in these difficult times. As a result, the FCA test case bore considerable importance. At the same time, the adoption of the test was also relatively risky given the untested nature of the Scheme. It is the success of the Scheme being used in the FCA test case which forms the centrepiece of this article. More specifically, this article has provided evidence to make a contemporaneous evaluation of the degree of success of the Scheme in the seminal FCA test case.

The first section introduced the Scheme and found that it is an expedited court process aimed at providing authoritative guidance before disputes arise. This made the FCA test case seemingly perfect for admission onto the Scheme. The second section evaluated the level of success the Scheme had in resolving each of the individual disputes. It found that the level of success varied across disputes. The third section assessed the overall level of success of the utilisation of the Scheme in the FCA test case and evaluated its potential for future and extended use. It found that the objectives behind the Scheme were largely met, whilst the objectives behind the FCA test case were met but not to the same degree. Immediately relevant authoritative English law guidance was provided, yet for some disputes, the degree to which the guidance serves to clarify contractual uncertainty is questionable. Notwithstanding these issues, it finds that the FCA test case was a good endorsement for the Scheme as it does away with many problems of the pre-existing methods. Thus, it is a positive development that the implementation of the Scheme onto the Financial List is permanent.

Overall, it can be stated that the endeavour was a success. The FCA test case, through utilising the Scheme, did provide immediate relevant authoritative English law guidance and, to a degree, this guidance subsequently clarified key issues of contractual uncertainty. It is certainly true to state that the FCA test case's success has been a good endorsement of the Scheme and has justified its implementation as a permanent fixture of the Financial List. While

the Scheme does require some improvements, these changes can be identified and implemented as future cases utilising the Scheme arise. The COVID-19 pandemic has had catastrophic consequences, but the FCA test case has helped in mitigating the potential fallout and chaos that could have occurred had the case not been conducted. The Scheme has not resolved all the issues stemming from the COVID-19 pandemic, but will likely have saved many people from financial ruin and kept many businesses operating. When viewed in this light, it is hard to say that the utilisation of the Scheme could be construed as anything other than a positive exercise, and it is likely that the aggrieved policyholders saved by this expedited process will agree.

Are the Laws Still Poor? Reflections on the Right to Work Limiting the Right to an Adequate Standard of Living

Chen First*

ABSTRACT

Despite the existence of a comprehensive human rights framework that includes the rights to work, social security, and an adequate standard of living, poverty continues to persist. This article posits that these rights, rather than being complementary, actually undermine one another and detract from discussions about human dignity with an individualistic interpretation. Louis Blanc's original conception of the right to work as a means of enabling the proletariat to shape the market and gain emancipation has been reduced to a focus on individual working conditions. The right to an adequate standard of living, which lacks concrete plans for implementation, is often conflated with the rights to social security and assistance at the national level. Thus, in most constitutions or court rulings, the right to an adequate standard of living is limited by placing the primary responsibility on the individual to work. Moreover, the current prevalent approach to addressing poverty is through the development of the job market. These cumulatively change the structure of the right to work to a duty to work, placing the main responsibility on individuals to alleviate their own poverty and blurring the distinction between state intervention and individual responsibility. This shift in focus from human dignity to work and working conditions also has negative societal consequences, as it tends to stigmatise those in poverty and ignores the root causes of impoverishment. Most importantly, with this human rights structure, the fundamental issue of human dignity is left aside, alongside the right to an adequate standard of living.

* LL.M London School of Economics and Political Science ('22), LL.B and Bachelor of Social Work, Hebrew University of Jerusalem ('19). The author thanks Dr Margot Salomon for her guidance.

INTRODUCTION

At the time of writing, 64,890 households are either homeless or threatened with homelessness in England alone.¹ It is futile to ponder the question of what the future might hold for these households as social mobility is still an unattainable dream for many, given that capital accumulation is becoming more centralised within a small share of the population. As Piketty shows, there is an increasing number of individuals from each generation who inherit wealth larger than what the least well-paid 50% of the population earn in a lifetime.² Often, individuals who have acquired wealth despite modest backgrounds are those who are most remembered, contributing to the assumption that this is a common occurrence. This phenomenon, known as ‘survival bias’ in behavioural economics, leads us to overestimate the prevalence of social mobility.³ The vicious poverty cycle does not release anyone who wants to escape: those in poverty start the ‘race of life’ at a disadvantaged position and face more obstacles than others.⁴ However, work has remained the onus of alleviating poverty.

The British Poor Relief Act 1601, known as ‘the Old Poor Laws’ (‘Poor Laws’),⁵ has been considered a focal point for the general ‘right to welfare relief’ by some scholars.⁶ Those laws were initially enacted as a response to the

¹ Department of Levelling Up, Housing & Communities ‘Statutory homelessness in England: October to December 2021’ (*Gov.uk*, 2022) <<https://www.gov.uk/government/statistics/statutory-homelessness-in-england-october-to-december-2021>> accessed 01 August 2022.

² Thomas Piketty, *Capital in the Twenty-First Century* (Arthur Goldhammer tr, The Belknap Press of Harvard University Press 2017) 420-421.

³ Lauren A. Rivera, *Pedigree: How Elite Students Get Elite Jobs* (Princeton University Press 2015).

⁴ OHCHR, ‘Report of the Special Rapporteur on extreme poverty and human rights’ (27 May 2015) UN Doc A/HRC/29/31, para 13.

⁵ British Poor Relief Act of 1601 (43 Elizabeth 1 c 4); The earliest English Poor Laws: the Statute of the Labourers 1350s; the Poor Law Reforms of the 1530s, the statute of Artificers of 1563; and the Elizabeth Poor Law of 1601 when taken together constitute ‘the main foundation of the English poor relief system’.

⁶ For example: D King, *Actively Seeking Work* (Chicago UP: London, 1995) 224 referring to the Poor Laws of 1601 including the principle of ‘deserving poor’ were adopted by the American State; William P. Quigley, ‘Backwards into the Future: How Welfare Changes in the Millenium Resemble English Poor Law of the Middle Ages’ (1998) 9 Stanford Law &

breakdown of feudalism that left many astray. Moreover, because the Black Plague and famine erased a third of the English population, gaps were created in the working market that needed to be filled. This led to a policy that forced non-workers to work.⁷ The Poor Laws ‘established the principle of a legal, compulsory, secular, national system of relief’, and gave England the reputation of a country where ‘compassion had become public policy’, in Himmelfarb’s words.⁸ These laws, though reformed from time to time, were retained for centuries until they were finally replaced by the welfare state after the Second World War. During that period, England served as a social laboratory.⁹

The three main pillars of the Poor Laws were: 1) a ‘poor rate’, a compulsory assessment in each parish financing outdoor relief for deserving indigent households; 2) a campaign against vagrants and beggars, in which begging and casual almsgiving were prohibited; and 3) an effort to provide work for the poor in order to reduce their need for relief and prevent them from wandering. These laws aimed to provide public benefits for the deserving poor and impose labour discipline.¹⁰

It appears that the historical significance of the Poor Laws is found in their spirit, which focuses on the individual’s behaviour and the importance of work as a means to alleviate poverty. The Poor Laws distinguished the ‘impotent’, deserving poor, i.e., victims of ill fate beyond their control, from the ‘able-bodied’, undeserving poor, who were only entitled to welfare under certain behavioural conditions. Hence, circumstances ‘beyond their control’ were the justification behind the assistance of the Poor Laws.¹¹ The Poor Laws also

Policy Review 101; Paul Spicier, *The Idea of Poverty* (Bristol University Press, Policy Press 2007) 73.

⁷ Quigley (n 6) 101, 103.

⁸ Gertrude Himmelfarb, ‘The idea of poverty’ (1984) 34(4) *History Today* 4.

⁹ *ibid* 5; In Europe between 1600 and 1900, social support took on the structure of Poor Laws, which attempted to increase work ethic among the poor and alleviate poverty; Jennifer Tooze, ‘The Rights to Social Security and Social Assistance: Towards an Analytical Framework 1’ in Mashood Baderin and Robert McCorquodale (eds), *Economic, Social, and Cultural Rights in Action* (Oxford University Press 2007).

¹⁰ Paul Slack, ‘Poverty in Elizabethan England’ (1984) 34(10) *History Today* 5-6.

¹¹ Amir Paz-Fuchs, *Welfare to Work: Conditional Rights in Social Policy* (Oxford University Press 2008) 66-73.

stipulated forms of punishment or coercion against the 'idle' poor,¹² such as imprisonment for disobedience.¹³ This led to the censuring of idleness, the rooting for production, and the labelling of distress as the individual's demerit.¹⁴ This structure placed the primary burden of alleviating poverty on the individual instead of examining it as a systemic societal problem.

Ideally, the following holistic human rights framework should have alleviated poverty: the right to work and earn a livelihood, the right to social security which protects from risks that prevent individuals from working, and the right to an adequate standard of living. Despite this framework, the impoverished simply continue to live in poverty.

This article aims to examine how the relationship between the right to an adequate standard of living and the right to work aligns with the rationale of the Poor Laws, which define poverty as an individualistic concern, scrutinising the individual's behaviour in relation to the labour market. The article will first introduce the concept of human rights and its legal framework to frame the discussion of rights and duties. Next, it will discuss the origins, the implementation, and the interactions between the aforementioned human rights, and examine how the right to an adequate standard of living is thereby undermined. Ultimately, the right to work is redefined as a duty to work, which places the individual as the primary stakeholder responsible for securing their own human rights. A change in discourse is observed, moving away from the concept of human dignity towards a focus on work and working conditions, which are perceived to be the main route to preserving human dignity and adequate standards of living.¹⁵

¹² Marjorie K. McIntosh, 'Poverty, Charity, and Coercion in Elizabethan England' (2005) 35(3) *The Journal of Interdisciplinary History* 457-459.

¹³ Quigley (n 6) 103.

¹⁴ Paul A. Fideler, 'Introduction: Impressions of a Century of Historiography' (2000) 32(3) *Albion* 381, 383.

¹⁵ Throughout the article, I will demonstrate the argument with the UDHR and the ICESCR as an example of the international agreement on the rights examined, acknowledging that it is not the whole list.

I. The concept of human rights and human rights law

The discussion will begin with a short analysis of the framework for human rights. Content-wise, rights can be viewed in two ways: according to the Interest ('Benefit') Theory of Rights or the Theory of Choice ('Will Theory'). In short, 'choice' theorists¹⁶ claim that the purpose of rights is to protect the control and autonomous choice of the individual.¹⁷ A right allows its holder the option of enforcing or waiving the correlative duty (the duty that third parties have with respect to the holder's right), which empowers the holder to be more autonomous in their choices.¹⁸ The 'Interest Theory' proclaims that a right exists if society has a corresponding duty to the right *and* this duty is rooted in the putative interest of the right-holder;¹⁹ that is, if the right benefits the individual. Thus, arguably, according to the Interest Theory, it is logical to restrict one's autonomy (or compel their actions) if it is for one's benefit. Hence, the way to turn a right into a duty is short.

Human rights can be summarised by the following formula: 'A has the right to B against C because of D'. The As are individuals protected by legal instruments, such as those under the jurisdiction of states that are signatories to that law. Most states have ratified the International Covenant on Economic, Social and Cultural Rights ('ICESCR'), and thus bear the duty of guaranteeing the three rights examined in this article. The justification for legal human rights adheres to the belief that being human inherently entails that one is born free and is equal in dignity and rights to others.²⁰ Hence, human rights are supposed to promote the interest and benefit of all human beings, regardless of socio-economic status, occupational status, or economic interests and without any attached duties or preconditions.

¹⁶ H.L.A Hart in England and Carl Wellman in the United States.

¹⁷ A. John Simmons, *The Lockean Theory of Rights* (Princeton University Press 2020) 93.

¹⁸ William A. Edmundson, *An Introduction to Rights* (Cambridge University Press 2004) 122.

¹⁹ Paz-Fuchs (n 11) 198; Paz-Fuchs, 'The Right to Work and the Duty to Work', in Virginia Mantouvalou (eds) *The Right to Work : Legal and Philosophical Perspectives*, (Bloomsbury Publishing Plc 2017) at 182.

²⁰ UNGA Res 217 A(III) (10 December 1948) A/RES/3/217 A.

Accordingly, B stands for the rights to which A is entitled.²¹ C is typically the government, as Article 2 of the ICESCR stipulates.²² Thus, the typical structure would be: one (A) has the right to work (B), against the state (C), because one is a human being (D).

In practice, rights embody three types of state duties: respect, protection, and fulfilment. The state should refrain from interfering with the enjoyment of a right and should protect human rights from breaches by third parties like private actors, and ensure that human rights are realised through the obligation on states to facilitate these rights.²³

Arguably, as Paz-Fuchs describes, assuming a right to X does not stipulate a duty to do X. Therefore, the right to speak does not impose the duty on one to speak.²⁴ Yet it appears that the right to work, by being perceived as beneficial to the individual (for example, because it allows one to participate in society and to lift oneself out of poverty) has turned the right into a duty that society expects one to undertake. This, to some extent, corresponds with the failure of the Theory of Interest since it is for the benefit of the individual to work.²⁵ Yet this precondition of working seems somewhat contradictory to the justification that human rights protect individuals by virtue of being born a human. This article will examine how the initial aim of the right to work corresponds with the Theory of Choice, which aims to expand the proletariat's autonomy. While the policy of poverty eradication aims to open working opportunities and enable the impoverished to seize employment, this however, seems to correspond more with the Theory of Interest. The prevailing policy of

²¹ Dan Seymour and Jonathan Pincus, 'Human Rights and Economics: The Conceptual Basis for Thier Complementarity' (2008) 26 Development Policy Review 387, 396-397; Human rights can be summarised by the following formula: 'A has the right to B against C because of D'. The As are individuals covered by legal instruments, for instance, those under the jurisdiction of states that are a signatory of that law.

²² While this paper acknowledges the perspectives that corporations and other third parties are subject to human rights obligations, this is beyond the scope of this paper.

²³ Walter Kälin and Jörg Künzli, *The Law of International Human Rights Protection* (2nd edn, Oxford University Press 2019), 87-89; International Commission of Jurists (ICJ), 'Maastricht Guidelines on Violations of Economic, Social and Cultural Rights' (UNHCR, 26 January 1997) <<https://www.refworld.org/docid/48abd5730.html>> accessed 2 April 2022.

²⁴ Paz-Fuchs (n 11) 193; Paz-Fuchs (n 19).

²⁵ *ibid.*

eradicating poverty works *mainly* through encouraging participation in the job market and providing social security when one cannot work. This turns work into a duty and a precondition to realising human rights, such as the right to an adequate standard of living.

II. Rights and Tensions

This section will examine the comprehensive framework of rights intended to address poverty: the right to work, the right to social security, and the right to an adequate standard of living. These rights and their interactions will be examined with a focus on the tension between ensuring that these rights are preserved on a large scale whilst being safeguarded at the individual level. This leads to several common criticisms of human rights, including that main right-holders are individuals.²⁶ Therefore, the primary objective of human rights is to address individual rights violations rather than their structural causes. Moyn claims that human rights are historically correlated with market fundamentalism since they are ‘unambitious in theory and ineffectual in practice’, reduced to the bare minimum in a capitalist economy that accepts inequality as long as there is a provision for minimum protection.²⁷ Human rights are supposed to counter trade liberalisation (an investment regime that mainly protects transnational corporations’ interests), spending cuts on welfare state regimes, deregulation, and the creation of ‘flexible’ work practices in favour of profit generation and privatisation.²⁸ With regard to inequality, Moyn emphasises the absence of a ceiling on the wealth gap between the rich and the poor in the Universal Declaration of Human Rights (‘UDHR’). This highlights the idealist nature of an equal distribution of human rights and challenges the notion that all human beings are born equal in dignity and rights.²⁹ This inequality is rooted in a power structure that deprives the impoverished of market opportunities and political voice.³⁰ Despite the fact that they may encounter questions related to

²⁶ Samuel Moyn, ‘Human Rights and the Age of Inequality’ in Doukje Lettinga and Lars van Troost (eds), *Can Human Rights Bring Social Justice?: Twelve Essays* (Amnesty International 2015) 17.

²⁷ *ibid* 16.

²⁸ John Linarelli, Margot Salomon and Muthucumarawamy Sornarajah, *The Misery of International Law: Confrontations with Injustice in the Global Economy* (Oxford University Press 2018) 226.

²⁹ *ibid* 13.

³⁰ Henry J. Steiner, Philip Alston and Ryan Goodman, *Is Poverty a Violation of Human Rights?* (Oxford University Press 2018) 309.

redistribution, human rights³¹ do not engage in pre-distribution issues³² and in the conditions that encourage the persistence of poverty.

The current discussions about providing a protection floor for *everyone* under international law³³ illustrate this point, demonstrating an understanding to settle for preventing or mitigating suffering without changing the terms that enable suffering.³⁴ The target is the impoverished individual, seen as the weak link in social protection, through intervention at the family-unit level.³⁵ Most importantly, the centralisation of the right to social security within the social protection floor scheme would support the main contention of this article, i.e., that *eventually* each individual is responsible for alleviating her own poverty. As part of the 2030 Agenda for Sustainable Development (SDGs), Goal 1 is to end poverty through 'social floors'.³⁶ This idea is elaborated in the International Labour Organisation's ('ILO') Recommendation No. 202,³⁷ which suggests that states should provide for basic income security for persons 'unable to earn sufficient income'.³⁸ This somewhat contrasts with the idea of social floors 'for all', even theoretically to those 'who opt to spend their life surfing waves'.³⁹ As Van Parijs explains, 'the extra leisure enjoyed by those unwilling to work would be stipulated as equivalent to the index of primary goods of the least advantaged'.⁴⁰

This tension between targeting the individual or the wider structure of inequality may be rooted in the fact that international law is mainly about rights and the corresponding obligations imposed on the state. Accordingly, states may adopt international law that protects or restricts certain economic activities, and establish a legal regime that corresponds to capitalism. While international law

³¹ This is applicable to International Law in general.

³² Linarelli, Salomon and Sornarajah (n 28) 35.

³³ *ibid* 251.

³⁴ *ibid* 36-37.

³⁵ *ibid* 252-253.

³⁶ UNGA, 'Transforming our world: the 2030 Agenda for Sustainable Development' (21 October 2015) A/RES/70/1, Goal 1.3.

³⁷ International Labour Organization Recommendation R202: Social protection floors (14 June 2012) (Geneva, 101st ILC session), part 2.

³⁸ *ibid* art 5(c); 5(b) for children; and 5(d) for elders.

³⁹ Philippe van Parijs, 'Why surfers should be fed: the liberal case for an unconditional basic income' (1991) 20(2) *Philosophy & Public Affairs* 101.

⁴⁰ *ibid*.

engages with state obligations, the liberal philosophy begins with the individual and their ability to create associations with others.⁴¹

1. The Right to Work

1.1. Origins

The discourse on the three rights begins with the most ancient among them, the right to work. In most countries, citizens satisfy their basic needs, such as shelter and food, through work.⁴² Inspired by John Locke, the post-revolutionary right to property was established on the idea of a ‘natural right’ to something acquired through labour, undermining corporate and public claims over individual entitlements.⁴³ In contrast, the French customary law of the Middle Ages allowed property to pass mainly through inheritance or marriage. Property signified the status of a family member, nobleman, or commoner. The main reform of the Revolution saw the right to property as separate and independent from caste affiliation.⁴⁴

Charles Fourier was the first to use the term ‘right to work’ in 1822 as substituting the natural right to hunt or gather food.⁴⁵ This term was circulated in French socialist literature and was a well-known maxim during the French Revolution. The right to work is rooted in the ‘right to existence’, as stipulated by Robespierre,⁴⁶ who posited that society is obliged to ensure the means of existence for its members - not as charity, but as a ‘duty of justice’. The underlying assumption was that the rights proclaimed in France in 1789 are universal. Additionally, it was also assumed that the enjoyment of some economic relief is necessary to enjoy these rights. Thus, a law that restricts a part of the population from a dignified subsistence (such as a law that perpetuates

⁴¹ Janet Dine and Andrew Fagan (eds), *Human Rights and Capitalism: A Multidisciplinary Perspective on Globalisation* (Corporations, Globalisation and the Law Series, 2006) 5.

⁴² Hugh Collins, ‘Is There a Human Right to Work?’ in Virginia Mantouvalou (eds), *The Right to Work: Legal and Philosophical Perspectives* (Hart Publishing 2015) 18.

⁴³ Dale K Van Kley (eds), *The French Idea of Freedom: The Old Regime and the Declaration of Rights of 1789* (Stanford University Press 1994) 302.

⁴⁴ Hannah Callaway, ‘*Revolutionizing Property: The Confiscation of Émigré Wealth in Paris and the Problem of Property in the French Revolution*’ (Harvard University 2015) 18-19.

⁴⁵ Pablo Scotto, ‘Thinking the Future of Work through the History of Right to Work Claims’ (2020) 46(8) *Philosophy & Social Criticism* 942, 944.

⁴⁶ *ibid* 945.

the predominance of property rights with unlimited freedom to trade) is contrary to the rights of man. The post-revolution right to work is a specific case of the right to existence, consisting of society's duty to provide sufficient work to everyone to have their basic needs fulfilled.⁴⁷

Yet, it was the socialist Louis Blanc, the father of the French constitutional right to work, who set the practical vision of the right to work as a part of a social reform rooted in political reform.⁴⁸ In Blanc's words, 'without a political reform, a social reform is not possible, for, if the second is the goal, the first is the means'.⁴⁹

The social reform concentrated on distribution and production, rather than integration into the capitalist form of work organisation. The focus was on the right to work as a means to bring about political change that would give power to the working class. The right to work is seen not only as an opportunity of employment, but as an opportunity for workers to take control independent of the private ownership of the means of production. In this regard, the state should ensure that it enables its workers to undergo skill development, associate with others, and participate in initiatives they most desire and are qualified to pursue. It is a right to free and associated work, rather than just minimum-wage labour, in contrast to the capitalist perception that allows the poorest to work for the benefit of the rich.⁵⁰

According to Blanc, the government should establish social workshops that would enable workers to pool their resources to purchase the tools needed for production. The profits generated by these workshops would be shared among the members of the association to support those who are unable to work, to help other industries facing difficulties, or to assist the people who may wish to join the association in the future. Overall, Blanc favoured the moderation of competition and promotion of solidarity:

⁴⁷ *ibid* 948.

⁴⁸ In this article I will refer mainly to Blanc's most fundamental work: organization of work (Louis Blanc, *Organisation du travail* (Bureau de la Société de l'industrie fraternelle 1847)) although he has further relevant works: Louis Blanc, *Le catéchisme des socialistes* (Aux Bureaux du Nouveau Monde 1849a); Louis Blanc, *Le socialisme: Droit au travail* (Aux Bureaux du Nouveau Monde 1849b).

⁴⁹ *ibid* 47.

⁵⁰ Scotto (n 45) 948.

In the system which we propose, crises would become rare. What causes them most frequently today? The veritable murderous contest between the interests, a contest from which no victor can come forth without leaving conquered ones on the field of battle.⁵¹

In this spirit of solidarity, Blanc claimed that there is a moral duty on each individual to contribute according to one's abilities:

The distribution of work and the distribution of its fruits would be based on ... the constitutive principle of the family: from each according to his faculties and to each according to his needs.⁵²

Thus, the Blancian conception of production entails everyone contributing freely, without hierarchical control, and according to their abilities. It remains a duty, but not one that the ruling class demands.⁵³

In terms of distribution, individuals would be equal and truly free if their basic needs were met and they could exercise their faculties.⁵⁴ The Blancian conception is about de-commodifying labour power - *not* by removing private property, but by universalising and democratising control of the means necessary to subsist.⁵⁵ Although the right to work is also somewhat a moral duty, it aims to liberate and empower workers to shape their own destinies. It promotes a culture of solidarity and acknowledges the inherent value and dignity of every individual, striving to align market conditions with these principles.

1.2. International legal instruments

An examination of the mid-20th century reveals an international conception of the right to work, where it is referred to more as an extension to

⁵¹ Blanc 1847 (n 48) 51-52.

⁵² Scotto (n 45) 950, referring to Louis Blanc, *Le catéchisme des socialistes* (Aux Bureaux du Nouveau Monde 1849), 28.

⁵³ *ibid.*

⁵⁴ *ibid* 949.

⁵⁵ *ibid* 953.

all decent-wage labour,⁵⁶ rooted in the experience of economic crises during the interwar period when unemployment was regarded as an important cause of support for the Nazi regime.⁵⁷ The right to work is stipulated in two international provisions: Article 23 of the UDHR, which implies mainly free access to the labour market, and Article 6 of the ICESCR, which is aimed at creating a well-functioning labour market to promote full employment.⁵⁸ Narrowly construed, the right to work under the UDHR seeks to ensure material survival and dignity as complementary to the right to life.⁵⁹ More broadly, it contributes to the self-development of the individual,⁶⁰ helping them adopt a social role within society and increase their self-esteem.⁶¹ The UDHR was created under a socialist influence, shaped by the communist bloc. Thus, many standard Latin American socialist constitutional provisions were adopted under the influence of the UDHR, including the right to work, although not according to Blanc's vision.

There are three main interpretations of the right to work. The right to work is firstly interpreted as a right to free choice of employment, referring to the right to choose one's occupation without unjustified discrimination or restrictions, which is a key right under liberal political theories adhering to justice standards based on equal opportunity.⁶² Throughout history, restrictions were often placed on refugees and migrants, ethnic minorities, and women.⁶³

⁵⁶ *ibid* 952.

⁵⁷ Kent Kallstrom and Asbjørn Eide, 'Article 23' in Gudmundur Alfredsson and Asbjørn Eide (eds), *The Universal Declaration of Human Rights: A Common Standard of Achievement* (Martinus Nijhoff Publishers 1999) 491.

⁵⁸ *ibid* 495.

⁵⁹ Commission on Human Rights Drafting Committee, Second Session, 29th Meeting, E/CN.4/AC.1/SR.29 (20 May 1948) in Ben Saul (ed), *The International Covenant on Economic, Social and Cultural Rights, Travaux Préparatoires 1948-1966* (1st edn, Oxford University Press 2016).

⁶⁰ UN Committee on Economic, Social and Cultural Rights (CESCR), 'General Comment No.18: The Right to Work (Art. 6 of the Covenant)' E/C.12/GC/18 (OHCHR, 2006) <<https://www.refworld.org/docid/4415453b4.html>> accessed 1 April 2022 para 1.

⁶¹ Preamble to Employment Promotion and Protection against Unemployment Convention (adopted 21 June 1988, entered into force 17 October 1991) C168.

⁶² Collins (n 42) 21.

⁶³ *ibid* 22.

This aspect of the right to work also implies ‘not being forced’ to work.⁶⁴ Economic necessity indeed forces people to work, yet it does not interfere with human liberty in the same way enslavement does.⁶⁵ This choice to work instrumentalises the value of work while respecting the individual’s dignity and freedom.⁶⁶ There was no extensive discussion about this aspect when drafting the UDHR, with the assumption that the individual is the author of their own life.⁶⁷ The idea of a ‘social obligation’ to work as a contribution to the general welfare of the state, based on the socialist belief that everyone has a responsibility to do work that is beneficial to society,⁶⁸ was rejected mainly due to Eleanor Roosevelt’s claim that it could lead to forced labour. Additionally, it was generally understood that there are times when it can be difficult to find work, such as after the Second World War when soldiers returned to domestic markets.⁶⁹ This decision seems wise since the duty to work cannot function without Blanc’s vision of a system of solidarity where everyone contributes but society is also obliged to meet everyone’s needs. The goal of satisfying the needs of all individuals in society will not be achieved by this measure alone as that might lead to forced labour.

The second interpretation focuses on the right to decent work, which is supposed to help one lead a dignified life. This interpretation refers to just and favourable conditions of work.⁷⁰ The ILO is responsible for the international labour standard.⁷¹ Currently, the standard entails a procedural method to ensure consent, engaging mainly with the following issues: freedom from forced labour; prohibition of child labour; freedom of association; access to jobs; strengthening social protection; social dialogue;⁷² and conditions of

⁶⁴ CESCR (n 60) para 6.

⁶⁵ Collins (n 42) 21.

⁶⁶ CESCR (n 60) para 4.

⁶⁷ Johannes Morsink, *The Universal Declaration of Human Rights: Origins, Drafting, and Intent* (University of Pennsylvania Press 2010).

⁶⁸ *ibid* 161-162. Examples include Brazil, Colombia, Costa Rica, Ecuador, Guatemala, Paraguay, and Uruguay, as well as the French constitution and some communist countries’ constitutions.

⁶⁹ *ibid*; Kallstrom and Eide (n 57) 493.

⁷⁰ Collins (n 42) 22-23.

⁷¹ International Labour Office (ed), *ILO Declaration on Fundamental Principles and Rights at Work* (ILO, 1998). Refer to the ILO’s follow-up reports for more information.

⁷² Guy Mundlak, ‘The Right to Work — The Value of Work’ in Daphne Barak-Erez and Aeyal M Gross (eds), *Exploring Social Rights: Between Theory and Practice* (Hart 2007) 348-349.

freedom, equity, security, and dignity.⁷³ Currently, the decency of work is attained through the SDGs, which tie together decent work and the capitalist aim of economic growth, as well as development-oriented policies.⁷⁴ A question arises here regarding the ‘just and favourable’ conditions meant to secure dignity. There seems to be a strong focus on free choice and remuneration in the ILO pillars, indicating a focus on material protection rather than a meaningful and fulfilling life as measured by other metrics. Both this aspect of the right to work, and the former aspect discussed, relate to the relationship between human rights and capitalism, as they are both grounded in the integration into the ‘free’ market, emphasising freedom and signalling the triumph over human enslavement.⁷⁵

Thirdly, the right to work also includes protection from unemployment, which has been criticised as vague. The implementation of social security schemes for the unemployed is redundant since other provisions in the UDHR cover this, such as Articles 22 and 25(1). This could mean restrictions on the basis of justified dismissals or compensation in the case of dismissals.⁷⁶

There is no right to be given work⁷⁷ because it forces the government to either act as an employer or to intervene heavily in the market to ensure that everyone is employed, which would likely be costly. Therefore, there is a balance to be struck between the public interest and individual rights. Under the UDHR, it was agreed that the obligation of the state is to reach full employment of its working population,⁷⁸ which arguably achieves the same result as directly providing work but without the accompanying costs. Under the ICESCR, the main state obligations are to prohibit compulsory labour, ensure equal access to decent work and training, and provide work for those that are unable to find it

⁷³ International Labour Organization, *The Rules of the Game: An Introduction to the Standards-Related Work of the International Labour Organization* (4th edn, International Labour Office 2019) 5.

⁷⁴ UNGA Res Transforming our world: the 2030 Agenda for Sustainable Development (25 September 2015) UN Doc A/RES/70/1 goal 8.

⁷⁵ Dine and Fagan (n 41) 4.

⁷⁶ Collins (n 42) 29-32.

⁷⁷ Kallstrom and Eide (n 57) 507; Saul (n 59); Commission on Human Rights Drafting Committee, Second Session, Australia: Draft of Additional Articles for the Draft International Covenant on Human Rights, E/CN.4/AC.1/21 (6 May 1948).

⁷⁸ Morsink (n 67) 162.

due to circumstances ‘beyond their control’. States must also implement policies to stimulate economic growth to raise standards of living and provide education and training programs.⁷⁹

Ultimately, the discussion revolves more around whether the individual *can* work and provide for themselves. Given that labour is used mainly for profit-making considerations,⁸⁰ the idea of ‘freedom to work’ in the capitalist framework is often used not to enhance the freedom of the worker to gain power, but to delegitimise effective tactics to improve working conditions. This obscures the urgency of collective action, which is another remnant of the socialist right to work stipulated under Article 23 of the UDHR. Arguably, this freedom to organise was also given a capitalist interpretation in order to balance the power dynamic between employers and employees, instead of allowing the workers to own the means of production. ‘Freedom’ is used to advocate for the individual freedom of the employee or employer to exercise their right to work according to their wish.⁸¹

In conclusion, the right to work changed from Blanc’s envisioned way of participating in the means of production and facilitating a political reform of solidarity, to the right to sell one’s own labour force (or, according to the neo-Marxist critique, the right to be exploited in a capitalist economy). It is an effort to enhance one’s right to control their own destiny at a basic level.⁸² Capitalist policy adopted the socialist right to work with an interpretation that corresponds to capitalist perceptions of individualism and liberalism, instead of upholding its original objective of empowering the proletariat to reform the market.

2. The Right to an Adequate Standard of Living

The atrocities committed during the Second World War increased the ubiquity of the term ‘human rights’. In 1944, President Roosevelt reaffirmed the need to codify economic and social rights in his famous speech on the ‘four

⁷⁹ OHCHR (n 4) para 23-26.

⁸⁰ Mundlak (n 72) 345-347.

⁸¹ Cedric de Leon, *The Origins of Right to Work: Antilabor Democracy in Nineteenth-Century Chicago* (ILR Press 2015) 2.

⁸² Morsink (n 67) 157-158.

freedoms’, including the ‘freedom from want’.⁸³ The message was that political rights are insufficient to realise true individual freedom, which cannot exist without economic security.⁸⁴ Human dignity is protected by this right, as it ensures that one is entitled to enjoy basic needs without being shamed, degraded, or deprived of basic freedoms, and that one is able to live above the poverty line.⁸⁵

2.1. The International Right to an Adequate Standard of Living

Roosevelt’s speech led the American Law Institute (‘ALI’) to convene a committee of worldwide experts and draft a proposal for an international bill of rights. Interestingly, this draft contained a similar set of rights discussed in this article: the right to work, food and housing, and social security.⁸⁶ The right to an adequate standard of living is stipulated under Article 25 of the UDHR and Article 11 of the ICESCR. The enjoyment of this right minimally requires the enjoyment of necessary subsistence rights, such as the right to own property⁸⁷ (which is not available to all and reveals the global inequality in distribution), the right to work, and the right to social security.⁸⁸

In relation to the UDHR, states were reluctant to support this right, agreeing on the general term ‘adequate standard of living’, while claiming that the list of issues (including housing or clothes) was too constricting. The specific

⁸³ Asbjørn Eide, Catarina Krause and Allan Rosas (eds), *Economic, Social, and Cultural Rights: A Textbook* (2nd edn, M Nijhoff Publishers 2001) 133-134; see also: United Nations Economic and Social Commission for Western Asia, ‘freedom from want’ (UNESWA, 2020)

<[⁸⁴ United Nations, Economic and Social Council \(1947\): Commission on Human rights, Drafting Committee, First Session, Summary record of eighth meeting, doc. E/CN.4/A.C.2/SR.8; Germana Aguiar Ribeiro Do Nascimento, ‘The Long Road to the International Recognition of Economic and Social Rights: The Right to an Adequate Standard of Living’ \(2018\) \(11\) *The Age of Human Rights Journal* 43, 45-46.](https://archive.unescwa.org/freedom-want#:~:text=In%20an%20address%20known%20as%20the%20Four%20 Freedoms,3.%20 Freedom%20from%20 want%2C%204.%20 Freedom%20from%20> accessed 5 March 2023.</p></div><div data-bbox=)

⁸⁵ Eide, Krause and Rosas (n 83) 133-134.

⁸⁶ Asbjørn Eide and Wenche Barth Eide, ‘Article 25’ in Gudmundur Alfredsson and Asbjørn Eide (eds), *The Universal Declaration of Human Rights: A Common Standard of Achievement* (Martinus Nijhoff Publishers 1999) 527.

⁸⁷ UNGA (n 20) art 17.

⁸⁸ Eide and Eide (n 86) 523.

list of issues ultimately prevailed because they were recognised as the most pressing issues globally.⁸⁹ Article 11 of the ICESCR has been elaborated on by the Committee on Economic, Social and Cultural Rights through several General Comments, which included the rights to adequate housing,⁹⁰ food,⁹¹ and water⁹² as its components.

There is no General Comment dedicated to the umbrella right of an adequate standard of living, thus leaving its practical meaning quite ambiguous. Neither the UDHR nor the ICESCR outlines specific methods for implementing this right,⁹³ perhaps to avoid a delay in its realisation.⁹⁴ Implementing this right through social security schemes was rejected,⁹⁵ with the acknowledgment that states have some discretion in choosing the means of implementation.⁹⁶

Examining the aforementioned General Comments on Article 11 of the ICESCR reveals two issues. Firstly, despite the fact that the state is encouraged to consider the wider framework of policies, legislation, and market conditions, states are given a wide degree of discretion. Arguably, the practical content of the right to an adequate standard of living is rooted in Article 2 of the ICESCR, which outlines the limits of state obligations to act ‘by all appropriate means, including particularly the adoption of legislative measures’.

⁸⁹ Morsink (n 67) 195-198.

⁹⁰ OHCHR, ‘CESCR General Comment No. 4: The Right to Adequate Housing (Art. 11 (1) of the Covenant)’ E/1992/23 (OHCHR, 1991) <<https://www.refworld.org/pd/fid/47a7079a1.pdf>> accessed 1 April 2022; OHCHR, ‘General Comment 7, The right to adequate housing (art. 11.1 of the Covenant): forced evictions’ E/1998/22 (OHCHR, 1997) <<https://www.refworld.org/docid/47a70799d.html>> accessed 1 April 2022.

⁹¹ OHCHR, ‘General Comment No. 12: The Right to Adequate Food (Art. 11)’ E/C.12/1999/5 (OHCHR, 1999) <<https://www.refworld.org/pd/fid/4538838c11.pdf>> accessed 1 April 2022.

⁹² OHCHR, ‘General Comment No. 15: The Right to Water (Arts. 11 and 12 of the Covenant)’ E/C.12/2002/11 (OHCHR, 2003) <<https://www.refworld.org/pd/fid/47a7079a1.pdf>> accessed 1 April 2022.

⁹³ Eide and Eide (n 86) 529.

⁹⁴ Saul (n 59); UNGA (11th Session) ‘Agenda Item 31, Draft International Covenants on Human Rights, Report of the Third Committee’ (9 February 1957) UN Doc A/3525.

⁹⁵ UNGA (n 94), UNGA Third Committee (11th Session), ‘Draft International Covenant on Human Rights, 739th Meeting’ (23 January 1957) UN Doc A/C.3/SR.739.

⁹⁶ Maastricht Guidelines on Violations of Economic, Social and Cultural Rights (n 23) para 8.

The second issue crucial to this article is the unclear balance between the responsibilities of the state and the individual in realising the right to an adequate standard of living. Where does the responsibility of the state end and the responsibility of the individual begin? There is an important personal responsibility on the individual to take all measures within their capacity to realise their right to an adequate life,⁹⁷ and that means realising what they are entitled to. Building on Amartya Sen's list of entitlement relations, this could be achieved by trade, for example, trading a possession with a willing party; production: producing goods using one's resources; work: acquiring through labour, which is similarly trade-based and production-based; and third parties: inheriting third-party entitlements through transfers and wills.⁹⁸

States have 'core obligations' to ensure the satisfaction of at least the minimum essential levels independent of its available resources, and a state in which any significant number of individuals is deprived of essential foodstuffs, basic shelter, and housing is, *prima facie*, failing to meet its obligations.⁹⁹ In any case, even if one is living in poverty, is this a *prima facie* claim, or did one simply not realise their entitlement? Uncertainties remain regarding the content of the right to an adequate standard of living, and a lack of discussion may enable states to interpret this right in line with the right to social security, a positive right to help an individual unable to work — emphasising the primary responsibility of the individual.

Even though it is not outlined in international law instruments, the literature provides insight into the duties of states under the right to an adequate standard of living. According to this literature, the level of adequacy under the right depends on the conditions in the society concerned.¹⁰⁰ To respect the right to an adequate standard of living, it is the duty of the state to enact policies that ensure that those entitlements can be realised. Noticeably, those entitlements are somewhat violated when the state itself generates injustice (and ultimately puts

⁹⁷ Eide, Krause and Rosas (n 83) 143.

⁹⁸ Amartya Kumar Sen, *Poverty and Famines: An Essay on Entitlement and Deprivation* (Oxford University Press 1992) 1-8.

⁹⁹ OHCHR, 'CESCR General Comment No. 3: The Nature of States Parties' Obligations (Art. 2, Para. 1, of the Covenant)' E/1991/23 (OHCHR, 1990)

<<https://www.refworld.org/pdfid/4538838e10.pdf>> accessed 1 April 2022, para 10.

¹⁰⁰ Eide, Krause & Rosas (n 83) 143.

the responsibility for alleviating poverty on the individual). For example, the lack of recognition of collective land rights for indigenous populations is a significant contributor to global poverty, particularly in North America, Australia, and New Zealand.¹⁰¹ Moreover, policies that prioritise the interests of urban elites (referred to as ‘development racism’) often disadvantage weaker groups and contribute to their economic hardship.¹⁰²

The obligation to protect entails, *inter alia*, the prevention of situations when people on the borderline of poverty cannot overcome crises and would have to sell their means. Under the obligation to fulfil, there is an obligation for the state to assist individuals through various methods such as training and price regulation,¹⁰³ and to provide opportunities.¹⁰⁴ This obligation is rooted in the justification that all other measures seem unproductive, especially in a market-oriented industrialisation. This can be done by measures of redistribution, such as progressive taxation.¹⁰⁵

It was emphasised that the right to work, which was meant to provide a basis for political and social reform, was rejected in this format and became individualistic. This section has argued that the right to an adequate standard of living allows for the same failure of fixation on the individual. The right to an adequate standard of living has the potential to transform market conditions and address the underlying causes of poverty through wide-ranging policies. By stipulating an approach centred around human dignity and addressing structural inequalities, the right to an adequate standard of living could lead to a discussion on how to achieve human dignity, particularly as to which institutions are concerned in promoting this cause, and how property rights might be adjusted to achieve human dignity, just as the right to work led to the right to receive property and income.

¹⁰¹ *ibid* 143.

¹⁰² Eide, Krause & Rosas (n 83) 143.

¹⁰³ *ibid* 144.

¹⁰⁴ Eide and Eide (n 86) 536.

¹⁰⁵ Eide, Krause and Rosas (n 83) 144.

3. Between the Right to Work and the Right to an Adequate Standard of Living: The Right to Social Security

The initial purpose of the right to social security is to support workers who have moved from the oppressive guardian of guilds in the feudal system to working ‘freely’, but with questionable freedoms of choice since their material conditions did not protect them from life risks. It was already understood more than a century ago that ‘a man who has no other source of income than his muscle or brain, if he loses one or the other, will not be saved from destitution by his freedom’.¹⁰⁶

The transition from an agrarian society to one characterised by wage labour during the industrial revolution brought about economic instability, particularly for those unable to support themselves.¹⁰⁷ The state, imbibed with the doctrine of liberalism, individualism, and *laissez-faire*, remained inactive until the community as a whole was threatened in its health and productivity by the suffering of its producers.¹⁰⁸ In 1883-84, Otto von Bismarck established the first social security system, serving as a model for modern social insurance. Here, employers paid contributions to finance workers during the periods when they could not work, later complemented by the introduction of old-age insurance.¹⁰⁹

The right to social security appears in Articles 22 and 25(1) of the UDHR covering various situations that result in a lack of livelihood, such as unemployment, sickness, or other circumstances ‘beyond one’s control’. Furthermore, Article 9 of the ICESCR provides for the right to social security, while Article 10 of the ICESCR stipulates the right concerning maternity leave. Additionally, the Social Security (Minimum Standards) Convention 1952 (No. 102) is an international treaty that establishes the basis for the development of social security systems worldwide. It covers nine areas of social security, including medical care, unemployment, old age, and invalidity. The Social Protection Floors Recommendation 2012 (No. 202) is a more recent policy document that provides a vision for achieving universal social protection in the 21st century. It recommends the establishment of national social protection

¹⁰⁶ International Labour Office, *General Problems of Social Insurance* (Issue 1 of Studies and Reports, Geneva International Labour Office 1925) 1.

¹⁰⁷ Martin Scheinin, ‘The Right to Social Security’ in Eide, Krause & Rosas (n 83) 211.

¹⁰⁸ International Labour Office (n 106).

¹⁰⁹ Scheinin (n 107).

floors, which are basic social security guarantees that would ensure access to essential healthcare and basic income security. These guarantees would include access to essential healthcare, basic income security for children, basic income security for working-age individuals unable to earn sufficient income, and basic income security for older persons. The ILO recognised nine branches of social security: medical care, sickness, invalidity, unemployment, employment injury, maternity, family, old age, and survivors.¹¹⁰

The legislated right to social security appears to be fulfilling its intended purpose (i.e., to support workers against life risks), but it is important to note that this right is limited in scope and serves the role of a fix for the failures in the capitalist market rather than a means of fundamentally transforming society and eliminating the structural issues that create poverty. Additionally, throughout the article, the theme of ‘beyond an individual’s control’ has been prevalent. There is a distinction similar to the one made in the Poor Laws between the ‘morally deserving’ poor and the ‘undeserving’ poor, with assistance being provided only when an individual is unable to provide for themselves due to moral circumstances. Thus, even Recommendation No. 202 which aims to provide *universal* coverage is restricted to the aforementioned categories. However, it would be naïve to believe that the impoverished are in control of their circumstances and are choosing to remain poor.

4. The Relationship between the Right to Social Security and Other Rights

4.1. The Right to Work

There is a complementary relationship between the right to work and the right to social security. The right to social security comes after the realisation of the right to work. The former right puts the *individual* as responsible for their standard of living by working, while the right to social security covers occasions of involuntary loss of income or an inability to provide for family necessities,

¹¹⁰ The ILO Social Security (Minimum Standards) Convention (adopted 28 June 1952) (No. 102). The European Code of Social Security 1962 enlarges the ILO recognition, requiring satisfying more branches accompanied by an Optional Protocol that provides for a higher level of benefits.

involving correlative duties on the beneficiary.¹¹¹ Social security has been a core element of the ILO's mandate, thus tying the two rights together while stressing that the best way to ensure income security is through decent work.¹¹² The ILO recently published that broader frameworks are needed for true universal protection, referring to taxes and social transfers, and to the Nordic countries that excel in that area with the lowest levels of inequality and the highest levels of happiness.¹¹³

4.2. An Adequate Standard of Living

The relationship between the right to an adequate standard of living and the right to social security is different on theoretical and practical levels, respectively. In theory, the international law instruments¹¹⁴ indicate a clear separation between the universal right to an adequate standard of living and the selective right to social security. Notably, while drafting the UDHR, this right to an adequate standard of living was merged into the same article as the one on social security, thus predicating one's entitlements upon being a member of a household with a breadwinner who is either employed, or unemployed for 'reasons beyond his control'.¹¹⁵ There was awareness of this issue among the delegations, as demonstrated by the amendment brought by New Zealand which made 'everyone' the only subject of the article. The right to social security, on the other hand, is selective since one is entitled to assistance only if they are unable to support themselves. The ILO argued that, by not attaching a list of covered cases to the right to social security, the meaning of this right would correlate with the meaning of the right to an adequate standard of living, which targets everyone. Ultimately, the difference was emphasised at the international

¹¹¹ As the Australian delegate explained during the drafting of the ICESCR in Saul (n 59) 1494; Commission on Human Rights (Sixth Session), 'Comments of Governments on the Draft International Covenant on Human Rights and Measures of Implementation' (22 March 1950) E/CN.4/353/Add.10.

¹¹² 'Social Security for All. Investing in Global Social and Economic Development. A Consultation. Issues in Social Protection; Discussion Paper 16' (*International Labour Organization*, 2006) <https://www.ilo.org/secsoc/information-resources/publications-and-tools/Discussionpapers/WCMS_SECSOC_8928/lang-en/index.htm> accessed 11 February 2023.

¹¹³ International Labour Office, 'World Social Protection Report 2020–22: Social Protection at the Crossroads – in Pursuit of a Better Future' (Geneva ILO, 2021) 31.

¹¹⁴ Referring to the UDHR and ICESCR.

¹¹⁵ Morsink (n 67) 198–202.

level, where it was stressed that the two rights have different aims and thus should be implemented differently and separately. However, as we will see, this difference was obscured at the state level.¹¹⁶

In the *travaux préparatoires* of the ICESCR, the right to an adequate standard of living was perceived as separate from the right to work. Initially, the USSR's phraseological suggestion of 'the right to work ... creating conditions which will remove the threat of death by hunger' was categorised under the right to an adequate standard of living.¹¹⁷ Later, it was decided that this right should be widened to protect the working conditions of the mother, *inter alia*, during periods of pregnancy and child rearing.¹¹⁸ The final article articulates a holistic approach, which includes the general principle of an adequate standard of living and the right to several essential needs,¹¹⁹ which extends to everyone.¹²⁰

In practice, the relationship between the right to an adequate standard of living and the right to social security appears to be one of a means to an end, where the right to social security is seen as a means to achieving an adequate standard of living. Thus, social security benefits are supposedly the main instrument to enable individuals to achieve an adequate standard of living.¹²¹ This seems somewhat unsuited to the aspired right to an adequate standard of living, which, as mentioned above, should be accomplished by promoting

¹¹⁶ *ibid.*

¹¹⁷ Saul (n 59); Commission on Human Rights (7th Session) 'Agenda Item 3(b), Working Group on Economic, Social and Cultural Rights, Compilation of Proposals Relating to Economic, Social and Cultural Rights' (27 April 1951) E/CN.4/AC.14/2Add.3.

¹¹⁸ Commission on Human Rights (7th Session) 'Agenda Item 3, Draft International Covenant on Human Rights and Measures of Implementation, 222nd Meeting' (8 June 1951) E/CN.4/SR.222.

¹¹⁹ Commission on Human Rights (9th Session) 'Draft International Covenants on Human Rights and Measures of Implementation, Economic and Social Council, Memorandum by the Secretary-General' (23 January 1953) E/CN.4/673; General Assembly Third Committee (11th Session), 'Draft International Covenant on Human Rights, 739th Meeting' (23 January 1957) A/C.3/SR.739.

¹²⁰ General Assembly Third Committee (11th Session), 'Draft International Covenant on Human Rights, 740th Meeting' (24 January 1957) A/C.3/SR.740.

¹²¹ OHCHR, 'General Comment No. 19 The right to social security (Art. 9 of the Covenant)' E/C.12/GC/19 (OHCHR, 2008)

<<https://www.refworld.org/docid/47b17b5b39c.html>> accessed 2 August 2022, para 22.

policies and legislation or any other method — not necessarily by addressing only those unable to work.

The redistributive character of social security is important for poverty reduction and alleviation, allocating financial assistance according to need and thereby promoting social inclusion.¹²² There is, however, a limitation on the scope of this effect due to the complementary relationship between the right to social security and the right to work, which positions the right to social security as an unwelcome intervention in the market. Therefore, rich states strive to demonstrate that they do not discourage individuals from working or undermine free market efficiency, while poor states are encouraged to privatise public services and prioritise free trade.¹²³ Most importantly, social protection bolsters political support for the same institutions and structures that facilitate poverty, as the state uses profits from capitalist growth economies to fund social security, thus legitimising economic growth efforts.¹²⁴

The next section will aim to demonstrate the symbiotic relationship between the right to an adequate standard of living and the right to social security, before analysing the primacy of the latter.

5. The Domestic Right to an Adequate Standard of Living is Limited

Theoretically separate from the right to work and social security is the right to an adequate standard of living. In practice, the following provisions illustrate states' conceptions about what those in poverty should be entitled to, and their constitutional protection of the citizen's wellbeing respectively. These

¹²² *ibid* para 3.

¹²³ Tooze (n 9) 331: 'Thus, structural reforms that are supported by organisations like the International Monetary Fund IMF and the World Bank to receive aid, generally speaking, intended to improve the efficiency and effectiveness of a country's economy by reducing government intervention in the economy. Some of these reforms may involve making the market more market-oriented, such as by reducing barriers to trade, improving the legal framework for businesses, and promoting competition'. See also: Vinaya Swaroop, 'World Bank's Experience with Structural Reforms for Growth and Development. MFM Discussion Paper; No. 11' (*World Bank*, 2016) <<https://openknowledge.worldbank.org/handle/10986/24360>> accessed 5 March 2023; Kyrgyz Republic, 'ESAF Policy Framework Paper 1998-2000' (*Kyrgyz ESAF*, 1998) <<http://www.imf.org/external/country/KGZ/index.htm>> accessed 5 March 2023.

¹²⁴ Linarelli, Salomon and Sornarajah (n 28) 252.

provisions mainly reflect the right to social security, allowing state assistance if one cannot work, rather than a general commitment to improve the standard of living or uphold human dignity.¹²⁵ It can therefore be argued that there is an unclear distinction between the right to work and social security, and the right to an adequate standard of living. This is perhaps because the means of implementing the right to an adequate standard of living is still undefined.

The following review of states' domestic law does not purport to be empirically comprehensive, but aims to demonstrate the muffled parameters delineating the right to work and the right to an adequate standard of living through two methods as laid out by Hirschl:

- (1) Contrasting the most different cases: when comparing along the axis of one specific variable, one should look for cases that differ on all other variables, in order to prove the consistency of the trend across the board.¹²⁶

As in our case, the tested variable is the right to an adequate standard of living in each jurisdiction's constitutional law. States were chosen based on their positions along the spectrum of capitalism and the degree of market intervention.¹²⁷ Additionally, there was an effort to include states representing a range of legal origins. These included English common law, which is focused on safeguarding citizens from government power; French civil law, which grants more authority to the government in managing citizens' lives and can be divided into French, Scandinavian, and German variants; and Socialist states, which are

¹²⁵ Avi Ben-Bassat and Momi Dahan, 'Social rights in the constitution and in practice' (2008) 36(1) *Journal of Comparative Economics* 103, 106.

¹²⁶ Ran Hirschl, 'On the Blurred Methodological Matrix of Comparative Constitutional Law' in Sujit Choudhry (eds), *The Migration of Constitutional Ideas* (1st edn, Cambridge University Press 2007) 51-56.

¹²⁷ See for example the following report measuring 'economic freedom' according to rule of law, government size and regulatory efficiency and open market indicators: The Heritage Foundation, '2023 Index of Economic Freedom' (*The Heritage Foundation*, 2023) <<https://www.heritage.org/index/ranking>> accessed 5 March 2023. Out of 177 states measured, the states in this report are in places: 1, 9, 10, 15, 43, 60, 153; The World Bank that measured 190 states, according to indicators that measure how easy it is to conduct business, and the states discussed within this article are in places: 2, 4, 20, 23, 35, 67, 108: World Bank, 'Ease of Doing Business rankings' (*The World Bank Group*, 2023) <<https://archive.doingbusiness.org/en/rankings>> accessed 5 March 2023.

rooted in the centralised regime of the Soviet Union.¹²⁸ It is assumed that all selected states aim to enhance the welfare of their citizens, and have therefore incorporated such a relevant provision in their constitutional laws. However, we will investigate whether this provision is constrained by the right to work. The conflicts between states' welfare provisions on one hand, and their enforcement of the right to work on the other, will help interrogate the widespread belief that individuals bear the primary responsibility for alleviating their poverty through working.

- (2) The most difficult cases: in order to verify a hypothesis, one should begin with the cases presenting the greatest difficulty, such that less difficult scenarios are proven *prima facie* by extension.

The assumption is that states addressing poverty via market intervention will not undermine the right to an adequate standard of living by instead placing the responsibility on the individual through the right to work.¹²⁹

Scandinavian countries are the most difficult cases, often characterised as adhering to values of universalism, solidarity (breaking down class division), and decommodification (the ability to live independently from the market since basic needs are provided or subsidised).¹³⁰ These countries are characterised by a 'social democratic consensus, facilitating the expansion of the welfare state'.¹³¹ However, Scandinavian social democrats are now making use of the welfare state to remain competitive in the global market, adopting targeted assistance instead of general welfare provisions due to concerns about people's incentive to work or care for each other.¹³²

Section 75 of Denmark's Constitution stipulates that:

¹²⁸ Avi Ben-Bassat and Momi Dahan, 'Social Rights in the Constitution and in Practice' (2008) 36 (1) *Journal of Comparative Economics* 11, 12, 103-119; the division method is from Thomas H. Reynolds and Arturo A. Flores, *Foreign Law: Current Sources of Codes and Basic Legislation in Jurisdictions of the World* (Rothman 1989).

¹²⁹ Ran Hirschl, 'On the Blurred Methodological Matrix of Comparative Constitutional Law' in Sujit Choudhry (ed), *The Migration of Constitutional Ideas* (1st edn, Cambridge University Press 2007) 51-56.

¹³⁰ Robert Cox, 'The Path-Dependency of an Idea: Why Scandinavian Welfare States Remain Distinct' (2004) 38 *Social Policy and Administration* 204, 205.

¹³¹ *ibid* 208.

¹³² *ibid* 212.

(1) to advance the public interest, efforts shall be made to guarantee work for every able-bodied citizen on terms that will secure his existence. (2) Any person unable to support himself or his family shall, when no other person is responsible for his or their maintenance, be entitled to receive public assistance ...

This provision is interpreted as stipulating reasonable efforts ensuring that a citizen who can work will be able to do so.¹³³ This provision does not create the basis to claim individual rights to employment. However, practically, state aid must be provided to those who cannot support themselves. A person must, however, exhaust her alternatives, including work or family support.¹³⁴ Ostensibly, the main way to achieve the right to an adequate standard of living in Denmark is tangled with the right to work. Roots of socialism are seen here with the emphasis on work for the benefit of the public interest as seen in many socialist constitutions in the past, consistent with Blanc's principle.¹³⁵ However, there is no redistributive element present in the right to work or the right to an adequate standard of living.

Similarly, Finland's Constitution demonstrates the classic structure of achieving human dignity, a structure whereby primary responsibility is placed on the individual:

Those who cannot obtain the means necessary for a life of dignity have the right to receive indispensable subsistence and care ... right to basic subsistence in the event of unemployment, illness.¹³⁶

¹³³ Susannah Pedersen, 'My Constitutional Act - My Constitutional Act with Explanations, 12th Edition' 44

<https://adsdatabase.ohchr.org/IssueLibrary/DENMARK_My%20Constitutional%20Act%20with%20explanations.pdf> accessed 6 July 2022.

¹³⁴ *ibid.*

¹³⁵ 'From each according to its faculties' Scotto (n 45) 950, referring to Blanc 1989 (n 50), 28. See also: Morsink (n 67), 161-162.

¹³⁶ s.19 Constitution of Finland (1999) (731/1999, amendments to 802/2007 included).

According to Finland's regulation, social assistance guarantees income security only as a last resort when personal or familial incomes are insufficient.¹³⁷ The term 'those who cannot obtain' implies defects in individual capabilities, and there is no reference to the state's responsibilities, again blurring the line of responsibility between the individual and the state.

The Colombian Court, with its roots in French civil law, developed the concept of '*minimo vital*', or the right of every individual to 'minimum condition for dignified life'. This right is derived from the right to life, the right to health, the right to work, and the right to social security.¹³⁸ It was ruled that in cases of extreme urgency where the basic needs of the individual are jeopardised, human dignity is violated, entitling the individual to sue for protection of her economic and social rights. The court has referred to this right in the context of salary, pension, maternity benefits, or benefits that are the only source of income for the individual.¹³⁹ Here, the right to an adequate standard of living is less restricted by the right to work as compared to previous jurisdictions, although it is still restricted to methods of income, and the primary responsibility of the individual is to exhaust them before qualifying for the state's welfare intervention, which is limited to cases of 'extreme urgency'.

Further meriting examination are countries with roots in German civil law, particularly Switzerland and South Korea. In Switzerland's Constitution,¹⁴⁰ reference to the wellbeing of the people can be found in Article 12, stipulating that '[those] *unable to provide* for themselves have the right to assistance and care, and to the financial means required for a decent standard of living'. Interestingly, Switzerland included a clause protecting human dignity under Article 7, yet criminalised begging regardless of this clause. In this regard, the European Court of Human Rights ruled that human dignity deems a person's means of

¹³⁷ European Commission, 'Your Social Security Rights in Finland' (*European Commission*, 2021) <<https://ec.europa.eu/social/BlobServlet?docId=13749&langId=en>> accessed 4 July 2022, 33.

¹³⁸ Magdalena Sepúlveda, 'Colombia: The Constitutional Court's Role in Addressing Social Injustice' in Malcolm Langford (ed), *Social Rights Jurisprudence: Emerging Trends in International and Comparative Law* (1st edn, Cambridge University Press 2009) 147-148.

¹³⁹ *ibid.*

¹⁴⁰ The Federal Constitution of the Swiss Federation (adopted 18 April 1999).

subsistence as a foundational component,¹⁴¹ thus positively recognising begging as a means to protect human dignity. The court ruled that human dignity protects the applicant's right to 'express her distress and to try and satisfy her needs by begging'.¹⁴² Still, a harsh position towards the impoverished, similar to the Poor Laws thus presented, prohibits begging but aids the eligible poor (those who are unable to provide for themselves).

The Constitution of South Korea¹⁴³ is another relevant example in contrast. On one hand, there is Article 32 which stipulates the right to work, and a relatively 'strong' duty to work on all citizens.¹⁴⁴ On the other hand, there is a 'weaker' duty on the side of the state to '*endeavour* to promote social security and welfare'.¹⁴⁵ Furthermore, this duty does not only sound non-committal, but was criticised in practice as being conceptually interpreted in relation to human dignity — which is an under-articulated concept in the court's jurisprudence — and also *de facto* void.¹⁴⁶ The right to welfare and the duty to work are not in the same article, but measuring the state's commitment to each could provide insight into the nature of the interactions among the individual, his rights, and the state's enforcement.

The case of Israel exemplifies the notion that constitutional law is not merely the written law, but rather a more dynamic phenomenon, formulated through the practice of law such as judicial decisions.¹⁴⁷ In Israel, the right to live with dignity has been recognised by the Supreme Court as a constitutional right. This includes access to basic necessities such as water, food and shelter and the 'duty of the state under the Basic Law: Human Dignity and Liberty ... to

¹⁴¹ Daniel Rietiker & Mary Levine, 'Begging the Question: *Lacatus v. Switzerland* and the European Court of Human Rights' Recognition of Begging as a Human Rights Issue' (2022) Harvard International Law Journal.

¹⁴² *ibid.*

¹⁴³ Constitution of the Republic of Korea (adopted July 17 1948, amended October 29 1987).

¹⁴⁴ *ibid* art 32: 'All citizens shall have the duty to work. The State shall prescribe by Act the extent and conditions of the duty to work in conformity with democratic principles.'

¹⁴⁵ *ibid* art 34(2).

¹⁴⁶ Chaihark Hahm, 'Constitutional Discourse on Human Dignity in South Korea: A Critical Appraisal' in Jimmy Chia-Shin Hsu (ed), *Human Dignity in Asia* (Cambridge University Press 2022) 62-86.

¹⁴⁷ Katharine G. Young, 'On What Matters in Comparative Constitutional Law: A Comment on Hirschl' (2016) 96 Boston University Law Review 1375.

maintain a system that will ensure a “protective net” for persons in society with limited means’.¹⁴⁸ In contrast, all state allowances are less than the minimum wage (reduced once one starts to earn income above a certain level) and evidence that one is seeking employment or has recently stopped working is required.¹⁴⁹

Additionally, concentrating on specific rights such as housing can arguably enforce the right to adequate standard of living without the working restriction. Singapore could shed some light here since they have *prima facie* solved the housing problem.¹⁵⁰ The state controls the land, construction, and sale of property. 82% of Singapore’s residents live in housing projects initiated by the state, and grants are provided for purchasing an apartment.¹⁵¹ However, despite being one of the richest countries in the world, Singapore’s poverty and inequality rates have largely been increasing since the 1970s, in spite of the absence of a formal poverty line.¹⁵² Nevertheless, the state’s main method of welfare assistance is formulated restrictively through the promotion of the right to work by deliberately lowering the quantum of welfare to around 5-8 percent

¹⁴⁸ HCJ 366/03 Commitment to Peace and Social Justice Society v. Minister of Finance [2005] (2) Israel Law Review 335.

¹⁴⁹ Without considering specific state allowances such as for the elderly or for children, Israel’s most significant allowances are those for the disabled and those for income support. See ‘Rates of income support benefit’ (*National Insurance Institute*, 2023) <<https://www.btl.gov.il/English/%20Homepage/Benefits/Income/%20Support/%20Benefit/Pages/Rates%20of%20Income%20Support%20Benefit.aspx>> accessed 12 March 2023; ‘Pension Rates’ (*National Insurance Institute*, 2023) <<https://www.btl.gov.il/English/%20Homepage/Benefits/Disability/%20Insurance/Pages/Rates%20of%20pension.aspx>> accessed 12 March 2023.

¹⁵⁰ See, for example, the UN Settlements Program awarding Singapore’s housing plan an award stating that ‘Singapore’s Housing and Development Board (HDB) will receive the award for providing one of Asia’s greenest, cleanest and most socially conscious housing programmes. For over half a century, HDB has housed a large slice of Singapore’s growing population. Currently, more than eight in 10 Singaporeans live in HDB apartments and more than nine in 10 own the apartment in which they live’: United Nations, ‘Innovative housing programme receives UN awards’ (*United Nations*, 2010) <<https://news.un.org/en/story/2010/09/350142>> accessed 3 January 2023.

¹⁵¹ Tan Shin Bin and Vignesh Louis Naidu, ‘Public Housing in Singapore: Examining Fundamental Shifts’ (2014) Lee Kuan Yew School of Public Policy, National University of Singapore 1, 4-6.

¹⁵² Mukul G Asher and Amarendu Nandy, ‘Singapore’s Policy Responses to Ageing, Inequality and Poverty: An Assessment’ (2008) 61 *International Social Security Review* 41, 46; Yen Kiat Chong & Irene YH Ng, ‘Constructing Poverty in Anti-Welfare Singapore’ (2017) 23(2) *Social Identities* 146, 147.

of per capita income, with the corresponding additions of income tax exemptions and income support targeting full-time workers.¹⁵³ This example illustrates how resolving one issue concerning impoverished people, such as housing, does not guarantee the elimination of poverty as a whole.

Notably, by making assistance schemes proportionate to income levels, as in Israel and Singapore, the principle of the Poor Laws' 'less eligibility' is illustrated. Under the Poor Laws, the criteria for allocating beds to non-working people, sometimes one to three, were significantly stricter in comparison to the allocative criteria for those who were working.¹⁵⁴ This comparison reveals another limitation of the right to an adequate standard of living by the right to work, namely relying on the latter's examination of one's occupational status as a basis for determining eligibility for the former.

The Canadian case of *Gosselin*¹⁵⁵ is another example of this. It concerned a challenge to a social assistance scheme that aids people aged under 30 with one third of the assistance that those above 30 receive, unless they participate in training or work experience employment programs. Interestingly, the court ruled that Section 45 of the Quebec Charter does not impose a sufficiently specific obligation compared to Article 11 of the ICESCR and Articles 22 and 25 of the UDHR, and is ambiguous enough to allow government discretion.¹⁵⁶ In other words, the court established certain limits on welfare assistance to encourage individuals to work. A criticism of Canada's ruling is that, by separating state power from the social and economic inequalities stemming from market forces, the court's rulings in these areas obscure the fact that markets can be directly influenced by government policy.¹⁵⁷ Moreover, the majority held that this restriction of social assistance does not impair the dignity of the young since the genuine objective to incentivise them to enter the working market is justified, irrespective of stereotypes.

It seems that there is a noticeable consensus about the existence of a right to some standard of living as a human right *that should be respected, protected or*

¹⁵³ Asher and Nandy (n 152) 56-58.

¹⁵⁴ Paz-Fuchs (n 11) 74.

¹⁵⁵ *Gosselin v. Quebec (Attorney General)* [2002] 4 S.C.R. 429.

¹⁵⁶ *ibid* [93], [420].

¹⁵⁷ Patrick Macklem, 'Social Rights in Canada' (2006) in Daphne Barak-Erez and Aeyal M Gross (eds), *Exploring Social Rights Between Theory and Practice* (Hart Publishing 2007).

provided by the state. Hence, most of the aforementioned jurisdictions include the right to ‘not be impoverished’ which is referred to as ‘life in dignity’ in Finland, ‘*minimo vital*’ in Columbia, the right to live with dignity in Israel, or ensuring the ‘right to existence’ in Denmark. However, these same jurisdictions still use the right to work as a limitation to one’s entitlement to welfare assistance, as enacted through written law, or by courts’ interpretation. While some countries such as Denmark and Finland (known as welfare states) stipulate this limitation explicitly in their constitution, other countries such as Singapore, Canada and Israel, incentivise individuals to work by referring to the right to an adequate standard of living as state assistance, followed by making the level of such assistance proportionate to the extent of individual employment. The significance of this review, however, lies in demonstrating that even tools of social reform built around the right to an adequate standard of living can fall under larger institutional structures perpetuating poverty when interpreted narrowly as a form of state assistance justifying the entitlement to social security.

To conclude briefly, it has been argued that, thus far, the right to work, rather than being a broad tool for social reform to empower the worker at large, has been too narrowly interpreted as mandating decent working conditions. In parallel, the right to an adequate standard of living was not substantially developed in the international realm and was interpreted as a complementary right to the right to work in domestic legislation, aiming to fill in the void when individuals cannot provide for themselves.

III. Implications: The Right to Work Turned Into a Duty to Work

Although the aforementioned jurisdictions do not stipulate the policy of ‘workfare’ per se, the logic behind it is demonstrated: the main route to alleviate poverty is through the employment market. Returning to Sen’s list of entitlements, it seems that an entitlement to property exists in the present context: the entitlement being state assistance when one cannot work. Assuming that impoverished people do not possess the right to an adequate standard of living, working or not being able to work and receive state assistance is the only way to alleviate their struggle. Working might not be a legal duty, but it is a social duty. Workfare can be defined broadly as ‘programmes or schemes that require people to work in return for social assistance benefits’. Thus, workfare entails the risk of a reduction in or loss of aid if an individual fails to comply

with working requirements. This attitude reflects policymakers' assumption that people dependent on assistance are disincentivised to enter the labour market and require a behavioural nudge to work.¹⁵⁸ Yet, at the same time, work itself is a necessary hurdle for the impoverished to overcome before receiving state aid,¹⁵⁹ especially when the former, serving as a prerequisite for the latter, can be inaccessible.

The last decade of the 20th century shifted the debate from levels of welfare expenditure to questions about its desirability and usefulness, casting doubt on its success in helping individuals reach self-sufficiency.¹⁶⁰ Since the Lisbon Summit of the European Union expressed its support for 'active labour market measures' which were widely adopted,¹⁶¹ its efforts primarily focused on training, education and national employment action plans, declaring that 'the best safeguard against social exclusion is a job'.¹⁶² International labour standards are referred to as essential for ensuring that the growth of the global economy provides benefits for all.¹⁶³ It is important to note that the SDGs' goal of reducing inequality by income growth of the bottom 40 percent of the population does not solve asymmetric wealth aggregation, leaving the wealth and power of the global 1 percent intact. Per some estimates, this inequality will take 207 years to eliminate at the cost of hastening catastrophic climate change, with our present production and consumption levels needing 3.4 Earths to sustain.¹⁶⁴

¹⁵⁸ Ivar Lødemel and Heather Trickey, 'A new contract for social assistance' in Ivar Lødemel (eds), *An Offer You Can't Refuse: Workfare in International Perspective* (Policy Press 2001) 6-8.

¹⁵⁹ Mundlak (n 72) 342.

¹⁶⁰ *ibid.*

¹⁶¹ Cox (n 130) 213.

¹⁶² Lisbon European Council, '23 and 24 March 2000 Presidency Conclusions' (*European Parliament*, 2000)
https://www.europarl.europa.eu/summits/lis1_en.htm#:~:text=The%20European%20Council%20held%20a%20special%20meeting%20on,social%20cohesion%20as%20part%20of%20a%20knowledge-based%20economy accessed 3 July 2023 [24]-[34].

¹⁶³ ILO (n 37) 7.

¹⁶⁴ Jason Hickel, 'The Problem with Saving the World' (*Jacobin*, 2015)
<https://jacobin.com/2015/08/global-poverty-climate-change-sdgs/> accessed on 31 January 2023.

The proposed UN Resolution to a Declaration on Human Rights Responsibilities is strikingly similar to socialist doctrines¹⁶⁵ in establishing the duty to work¹⁶⁶ and to participate as stipulated. This is important insofar as socialism tends to be conceived of as an ideology that aims to strengthen the working class. When considered independently from other relevant rights, however, the duty to work actually results in the opposite outcome. The working class, on whom duties to work are imposed by socialism without corresponding entitlement to rights such as the ownership of the means of production, become disempowered in practice. Today, this duty is stipulated in Article 29(6) of the 1981 African Charter of Human and People's Rights, and in some countries' constitutions, such as Article 35 of the Constitución Española in Spain, and Article 42 of the Constitution of the People's Republic of China.¹⁶⁷

There are several shared rationales between the Poor Laws and the idea of workfare: (1) social control: the public, who pay for the state's welfare programs, expect the impoverished to increase productivity as the public defines it (i.e. working); (2) the efficient use of resources, monetary and human: the principle of 'less eligibility' arises from a micro perspective, assuming rational, unemployed individuals would not work if the provision of benefits had remained above a certain level — on a macro level, human resources should be utilised to increase society's productivity and wealth; (3) social contract: *quid pro quo*, and 'something for nothing' is immoral and unjust.¹⁶⁸ The argument that incentives are needed to encourage people to work and become self-sufficient, rather than relying on state assistance, is problematic if this discourse overshadows discussions about social injustice and the actions that the state can take to combat poverty. This discourse can in fact also perpetuate harmful stereotypes about the laziness of impoverished individuals.

To conclude briefly, the right to work, which was initially aimed at being an opportunity for the disaffected to flourish, is now a precondition for gaining access to state assistance. However, it is possible that the ultimate loss for the proletariat is the conceptualisation of the right to work as a tool for

¹⁶⁵ Paz-Fuchs (n 19) 182-183.

¹⁶⁶ UN Economic and Social Council Official Records (55th Session), 'UN Resolution to a Declaration on Human Rights Responsibilities' (1999) UN Doc A/RES/53/144.

¹⁶⁷ Constitution of the People's Republic of China (adopted December 4 1982).

¹⁶⁸ Paz-Fuchs (n 11) 80-89.

reform, along with the narrow interpretation of an adequate standard of living in forms such as social security, as supplementary to the right to work.

IV. Reconsidering the concept of human rights law

1. The individual as the primary responsible actor

Conceptualising the right to work as a duty shifts the individual from being the right holder into being the duty holder. There are practical and social implications for the individual in this regard.

In practical terms, this raises the question of when an individual's duty to work ends and their right to an adequate standard of living begins, as well as the fate of this relationship. It could be argued that these rights occur simultaneously, with individuals required to work while the state protects their right to an adequate standard of living by regulating the market, among other measures, and providing benefits such as state assistance. Thus, state assistance which actively imposes an individual duty to fulfil the right to an adequate standard of living only becomes relevant if people are unable to satisfy their needs on their own,¹⁶⁹ and there is no substantive reference to this right otherwise. Yet, there is also a core obligation that assumes *prima facie* violation of the right to an adequate standard of living that might shift the primary responsibility from the individual to the state. This presupposes the logically prior responsibility of the state to take steps to alleviate poverty without qualification, by requiring the individual to have exhausted her right to work beforehand.¹⁷⁰ The lack of specific CESCR Comments on the general principle of an adequate standard of living makes it difficult to determine what this core obligation is, irrespective of the preliminary assumption that this obligation is violated when people are living in poverty. Assuming the idea that most states violate human rights might be a radical one, but reframing the narrative on poverty and rights in this way nonetheless could lead states to better protect their citizens' rights. This approach challenges the common stereotypes plaguing the impoverished, and emphasises the state's responsibility to ensure human rights. This issue becomes more severe considering the lack of universally-accepted criteria as to what individual entitlement to the right to

¹⁶⁹ Seymour and Pincus (n 21) 88-89.

¹⁷⁰ Comment No. 3, (n 99) para 10.

work entails, and what specific obligations fall on public authorities.¹⁷¹ Thus, many impoverished individuals continue to drown in the sea of ambiguity, where they live in poverty despite being theoretically entitled to many rights.

From a social standpoint, the decision to structure the right to work as a duty represents a lapse in the moral discourse about the poor. By recognising the impoverished as rights holders, the discourse shifts to the basic structure of human rights: A is entitled to B by C, with the primary question being – is this right being realised, i.e., has C ensured that A has realised B? Hence, poverty can be perceived as a failure on the part of the state or as a systemic issue rather than an immoral or functional failure of the individual.¹⁷² A rights-based approach allocates entitlements and empowers the eligible to realise their needs. Seemingly, constructing the right to work as a duty stymies this progress. It potentially risks a return to a needs-based approach that problematically construes the delivery of services to the marginalised as dependent on goodwill and charity by the powerful.

Throughout history, there has been discourse over who is responsible for the impoverished. The right to work was absent in the first post-French Revolution constitution, partly because the discussion about allocating jobs was overly interspersed with the right to government assistance.¹⁷³ Years later, Mr Shann, the Australian delegate writing the ICESCR, addressed the concern that the state's commitment to promoting welfare would discourage people from working:

There was no reason to fear that such conditions would kill the spirit of initiative, for man was naturally ambitious; yet despite that natural ambition it was essential to protect the less fortunate and the weaker.¹⁷⁴

Mr Shann presented an important argument that portrays the impoverished in a more uplifting fashion. In essence, 'they' [the impoverished] do not need help since they are lazy, and thus 'we' [society] must activate their

¹⁷¹ Katarina Tomaševski, 'Indicators' in Eide, Krause & Rosas (n 83) 536.

¹⁷² Christine Chinkin, 'The United Nations Decade for the Elimination of Poverty: What Role for International Law?' (2001) 54 *Current Legal Problems* 553, 564-565.

¹⁷³ Scotto (n 45) 948.

¹⁷⁴ Saul (n 59).

moral agency and encourage them to work. Instead, being socioeconomically disaffected is a matter of being less fortunate than others: sheer luck *beyond an individual's control*. Poverty is inherently random and arbitrary.

This discourse focuses primarily on the idea of the welfare state and the question of who is responsible for the impoverished. However, it seems that this discourse is irrelevant. Firstly, there is a misconception of the welfare state as the benefactor of the impoverished.¹⁷⁵ In contrast, the welfare state prioritises social insurance and social rights, rendering the chief beneficiaries the middle class and the employed.¹⁷⁶ For example, pensions are generally reserved for the more affluent in some countries, unavailable for the poor.¹⁷⁷ Most importantly, this discourse is rooted in the relationship between the capitalist market and the welfare state, which is functionally necessary for capitalism to solve market failures but structurally incompatible. Here, privately determined economic action and publicly determined social protection are ‘shackled together’ while each structure works to simultaneously sustain but undermine the other. Capitalism assumes that the expansion of *laissez-faire* logic will ultimately result in a greater improvement in welfare,¹⁷⁸ while redistributions that modify the primary distribution based on property rights and market transactions will be inefficient by definition, and lower in priority.¹⁷⁹ Moreover, the idea of welfare did not grow in a void. Instead, it was initially conceived as a means of mitigating the effects of market failures and hardship.¹⁸⁰ By the 1960s, every developed nation had, at their core, elements of a welfare state — even the USA reacted to the Great Depression with the New Deal.¹⁸¹ This imagined dichotomy between the welfare state and capitalism often overlooks other alternatives that are not based on capitalism, since in practice the welfare state is itself a form of capitalism insofar as its basis lies in improving the market’s allocative efficiency through intervention. Additionally, it fails to consider the

¹⁷⁵ David Garland, ‘The emergence of the idea of “the welfare state” in British political discourse’ (2021) 35(1) *History of the Human Sciences* 132, 139.

¹⁷⁶ *ibid* 135-136.

¹⁷⁷ Paul Spicker, *The Idea of Poverty* (Bristol University Press 2007) 78.

¹⁷⁸ Paul O’Connell, ‘On Reconciling Irreconcilables: Neo-Liberal Globalisation and Human Rights’ (2007) 7 *Human Rights Law Review* 483, 487.

¹⁷⁹ David Garland, *The Welfare State a Very Short Introduction* (Oxford University Press 2016) 51-53.

¹⁸⁰ Garland (n 175) 143.

¹⁸¹ *ibid* 150.

possibility of altering the pre-distribution conditions, as it primarily focuses on redistribution.

The focus on individual failings, when construing the impoverished, emphasises the personal characteristics of those excluded and overshadows the structural issues that have created and perpetuated poverty. This creates a lack of empathy, stigmatisation, and dehumanisation of the unemployed.¹⁸² Establishing the right to work merely demands individual personal responsibility over structural circumstances largely outside their control, and absolves the state of its obligation to the socioeconomically disaffected.¹⁸³ Poverty is a complex, multi-sectoral issue that cannot be solved through simple solutions.¹⁸⁴ It is a structural phenomenon that should be tackled as a matter of government priority.¹⁸⁵

Placing the duty to secure livelihood mainly on the individual seems unproblematic and ‘neutral’, but ignores the bigger picture. The economic literature finds substantial evidence of discrimination and structural pay gaps. For example, in Bangladesh, researchers found that poverty arises from a lack of opportunities as opposed to intrinsic, personal characteristics.¹⁸⁶ Additionally, it was found that the poverty trap creates a mismatch between talent and jobs — this misallocation of labour is rife amongst the impoverished.¹⁸⁷ In Israel, workers are three to four times more likely to find employment where a past co-worker of the parent currently works than in otherwise similar firms.¹⁸⁸ In Canada, 16 percent of resumes sent with English-sounding names, featuring Canadian education and experience, received a call back from an employer,

¹⁸² Paz-Fuchs (n 11) 189.

¹⁸³ Mundlak (n 72) 351.

¹⁸⁴ UN Committee on Economic, Social and Cultural Rights, ‘Substantive Issues Arising in the Implementation of the International Covenant on Economic, Social and Cultural Rights: Poverty and the International Covenant on Economic, Social and Cultural Rights’ (4 May 2001) UN Doc E/C.12/2001/10 [9].

¹⁸⁵ Tomafevski (n 171) 537.

¹⁸⁶ Clare Balboni and others, ‘Why Do People Stay Poor?’ (2022) 137(2) *The Quarterly Journal of Economics*, 785.

¹⁸⁷ *ibid.*

¹⁸⁸ Shmuel San, ‘The Role of Social Connections in the Labor Market’ (*The Hebrew University of Jerusalem* 2022) <https://mulysan.github.io/San_JMP.pdf> accessed 5 March 2023.

compared to only 5 per cent for resumes with foreign-sounding names.¹⁸⁹ In the USA, a similar effect was found for African-Americans with Black-sounding names compared to those with Caucasian-sounding names.¹⁹⁰

Framing the right to work as a duty blurs the line between the individual and state, raising questions about whether the individual has done enough to avoid poverty and when the state should intervene. This becomes especially important when there are no clear legal methods by which to implement the right to an adequate standard of living. As a result, the individual may be stigmatised for not working, and thus further marginalised from society, while the larger social and economic systems that create and maintain poverty are not addressed. This individualistic approach to addressing poverty through human rights may be insufficient in addressing the root causes of poverty.

2. The Entitlement: Work and the Disappearance of Human Dignity

Although the duty to work does not stipulate an entitlement to demand work, it puts ‘working’ at its centre. Any discussion of the right to an adequate standard of living separate from the right to work might be ignored if the impoverished are first asked *if* they can work. This can result in a narrow definition of work as a means to alleviate poverty, and hence as a means of survival, rather than it being a means to develop one’s personality, self-esteem, or social role. Consequently, interpreting human dignity based on working conditions may lead to a narrower interpretation of the concept based on living conditions as a basic matter of survival.

Work-related rights generally only safeguard the ability to work,¹⁹¹ working conditions, the meaning of decent work, or social coverage when one cannot work. In contrast, the right to an adequate standard of living puts the

¹⁸⁹ Philip Oreopoulos, ‘Why Do Skilled Immigrants Struggle in the Labor Market? A Field Experiment With Six Thousand Resumes’ (2009) National Bureau of Economic Research working Paper 15036, 24
<<https://www.nber.org/digest/sep09/why-do-skilled-immigrants-struggle-labor-market>> accessed 30 July 2022.

¹⁹⁰ Roland G. Fryer and Steven D. Levitt, ‘The Causes and Consequences of Distinctively Black Names’ (2004) 119 (3) *The Quarterly Journal of Economics* 767, 771.

¹⁹¹ Krzysztof Drzewicki, ‘The Right to Work and Rights in Work, in Economic, Social and Cultural Rights’ in Eide, Krause & Rosas (n 83) 223-224.

individual at the centre of analysis and means, *inter alia*, that one can enjoy basic needs under conditions of dignity. Accordingly, human dignity should be achieved without degrading or denying one's freedoms to satisfy one's needs.¹⁹²

The employed and the unemployed are two sides of the same coin in our context. Regarding those that do not work, there is a preliminary question about the definition of working and thus the extent to which non-workers are considered to be workers and the level of protection afforded by international conventions. Domestic workers, volunteers, and unpaid interns, have remained outside the ILO Conventions and are thus not covered by social security schemes for example.¹⁹³ Domestic work is excluded from these Conventions on the condition that an obvious condition to 'decent' work is a salary. This disproportionately affects women since they perform most of the domestic work across all cultures, age groups and social classes.¹⁹⁴ Domestic work entails costs, and foregone opportunities to engage in paid work, study, or simple enjoyment of leisure time.¹⁹⁵ There is also uncertainty regarding people who do not work because of circumstances that are not recognised as justified under human rights law or the right to social security, such as those who are unable to work because they suffer from emotional difficulties or are the primary caregiver for a family member. Such portions of the demographic are not covered by international conventions either.

In regard to the employed, employment does not necessarily promise human dignity.¹⁹⁶ Considering work as a means of alleviating poverty may revert the conversation to the narrow view of work as an economic necessity that ensures human dignity. According to this argument, it is more detrimental to one's autonomy or self-esteem to beg for assistance as opposed to work, which may be unpleasant, but most people prefer it so that they can meet their essential needs and be treated with dignity. Nickel talked about the 'humiliating trilemma of starvation, a life of crime, or dependence' when one does not

¹⁹² Eide, Krause & Rosas (n 83) 133.

¹⁹³ UN Committee on Economic, Social and Cultural Rights, 'Right to just and favourable conditions of work (Article 7 of the International Covenant on Economic, Social and Cultural Rights) Draft' (20 January 2015) UN Doc E/C.12/54/R.2.

¹⁹⁴ Shahra Razavi, 'The 2030 Agenda: Challenges of Implementation to Attain Gender Equality and Women's Rights' (2016) 24(1) Gender & Development 25, 31.

¹⁹⁵ *ibid* 31.

¹⁹⁶ Collins (n 42) 29-32.

work.¹⁹⁷ However, the worth of life extends beyond the liberty to work.¹⁹⁸ More broadly, the pursuit of work is not simply economic in nature — ‘all human beings ... have the right to pursue ... spiritual development in conditions of freedom and dignity’.¹⁹⁹ Work engenders the development of human competencies, enhancing self-confidence. Conversely, a perceived lack of meaning in one’s employment diminishes cognitive capabilities and self-worth.²⁰⁰ As the sociologist E. Perry explains:

When a garbage-man feels stigmatised by the work he does ... the stigma shows in his eyes ... to avoid contaminating us with his lowly self. He looks away; and we do too. Our eyes do not meet. He becomes a non-person.²⁰¹

It is questionable whether those workers possess an adequate standard of living.

Though beyond the scope of this article, it is questionable if labour rights tackle economic insecurity and effectively respond to the deficiencies of the global labour market.²⁰² This question is more relevant considering that labour rights are influenced by the IMF and the World Bank, whose liberalisation measures have reduced the ability of states to secure work-related rights,²⁰³ and most of the world's poor are employed but have low earnings.²⁰⁴

Therefore, to genuinely affirm human dignity, it should be considered separately from the right to work, which could enforce accountability on the part of the state. The German Federal Constitutional Court (*Bundesverfassungsgericht*) ruling on the case of ‘Hartz IV Legislation’²⁰⁵

¹⁹⁷ *ibid* 18-19.

¹⁹⁸ Mundlak (n 72) 344.

¹⁹⁹ Article 1(a), 2(a) ILO Declaration of Philadelphia.

²⁰⁰ Andrea Veltman, ‘Is Meaningful Work Available to All People?’ (2015) 41(7) *Philosophy & Social Criticism* 725, 726-727.

²⁰¹ Stewart E Perry, *San Francisco Scavengers: Dirty Work and the Pride of Ownership* (University of California Press 1978) 7.

²⁰² OHCHR (n 4) para 13.

²⁰³ Kallstrom and Eide (n 57) 508.

²⁰⁴ Balboni and others (n 186) 786.

²⁰⁵ Bundesverfassungsgericht [BVerfG] [Federal Constitutional Court] Feb. 9, 2010, 175 *Entscheidungen Des Bundesverfassungsgerichts* [BVERFGE] (Hartz IV) 1, 125 (Ger).

demonstrates this point by establishing a constitutional right to a 'subsistence minimum' based on the declaration of human dignity in Article 1(1) of the Basic Law in conjunction with the Social Welfare State Principle in Article 20(1).²⁰⁶ The court found that the provisions relating to standard benefits are unconstitutional due to the lack of underlying statistical research by the legislature that proves their adequacy to human dignity,²⁰⁷ and that the legislator should respect and positively protect human dignity.²⁰⁸ The new right cannot provide quantifiable requirements, but it does require a review of the basis, calculations, and method of the assessment of benefits to ensure that they do justice to the fundamental right to dignity.²⁰⁹ The court found that the procedure used to determine the benefits violated the constitution.²¹⁰ Notably, the court found that the minimum standard is not as minimal as one would expect. It includes physical survival (food and water), minimum participation in human interaction (covering phone bills for example), and participation in cultural and social life (subsidising sports clubs and going to the cinema), although as we get further from the core right more discretion is afforded to the state.²¹¹ The court even ruled that every expense that is reduced should be justified by data.²¹² Furthermore, benefits are calculated not only by reference to net income, like the principle of 'less eligibility', but rather according to 'consumption and the cost of living'.²¹³ Consequently, it does not matter whether one is employed, and indeed the question does not arise; neither does the link between dignity and employment. A simple objective of the inquiry is to determine what one needs to live a dignified life.

The court used the so-called 'Object Theory' (*Objektformel*), which stipulates that human dignity prohibits treating human beings as mere objects.

²⁰⁶ Grundgesetz für die Bundesrepublik Deutschland (adopted 23 May 1949).

²⁰⁷ *ibid* [137]-[146], [220].

²⁰⁸ *ibid* [133]-[134].

²⁰⁹ *ibid* [142].

²¹⁰ Stefanie Egidy, 'Casenote — The Fundamental Right to the Guarantee of a Subsistence Minimum in the Hartz IV Decision of the German Federal Constitutional Court' (2011) 12(11) German Law Journal 1961, 1962.

²¹¹ Claudia Bittner, 'Casenote — Human Dignity as a Matter of Legislative Consistency in an Ideal World: The Fundamental Right to Guarantee a Subsistence Minimum in the German Federal Constitutional Court's Judgment of 9 February 2010' (2011) 12 German Law Journal, 1941, 1952; Hartz IV (n 205) [135], [138].

²¹² Hartz IV (n 205) [175].

²¹³ *ibid*, [183]-[187].

There is no need to examine one's achievements or social status to guarantee entitlement to human dignity.²¹⁴ According to this principle, individuals who do not follow certain societal norms or expectations are not stigmatised and are seen as equal to others. The ruling focuses on societal conduct affirming individual dignity by virtue of being human instead of providing resources to support individual participation in the workforce, although the latter is not redundant.

Generally, the court frames 'human dignity' as distinct from social status. Through this, the state does not 'give human dignity to the poor', but instead merely enables 'every individual to lead a life of self-determination and autonomy'.²¹⁵ The court ruling was labelled as an example of the 'Communitive Theory': the idea that human dignity is a relational and communicative concept derived from social recognition, which promotes intersubjectivity and is closely tied to one's sense of belonging within a national community.²¹⁶ The right to work was similarly conceptualised by Blanc as active participation in a society that was intersubjective. Moreover, it was ruled that benefits cannot be denied, even if the person in need is responsible for creating that need.²¹⁷ It is interesting to note that throughout this article, the right to work was seen as a right that enables one to participate in society, attain a social role, and promote self-development. In this ruling, the court demonstrated that one does not have to earn their place in society, but rather that the right to an adequate standard of living opens the gates to society simply by virtue of one's human dignity. The state does not allocate charity, but rather allows all to live dignified and autonomous lives. Human dignity is not earned; it is axiomatic.

The court thus turned human dignity into a socio-economic right belonging to every person.²¹⁸ Resolving the issue of socio-economic rights, as well as the accountability concerns that arise alongside them, is important to the right to an adequate standard of living. For example, it was already recommended in general to use the term 'violation' of socio-economic rights only when a legal obligation exists and failure to uphold it can be proven. There are many bleak economic situations, and there is a reluctance to undermine the

²¹⁴ Egidy (n 210) 1969.

²¹⁵ *ibid* 1970.

²¹⁶ Bittner (n 211).

²¹⁷ Egidy (n 210), 1972.

²¹⁸ *ibid*, 1951.

seriousness of allegations about rights violations.²¹⁹ However, the right to an adequate standard of living remains a vague obligation in international law, while poverty is a multi-sectoral issue without a singular cause and solution.²²⁰ Socio-economic rights are already conceived of as secondary to civil and political rights, since they place a financial burden on the state.²²¹ However, this separation between socio-economic rights on the one hand and civil and political rights on the other is artificial, since it was also recognized that infringement of socio-economic rights affects the civil and political rights of, for example, the homeless that suffer from a lack of privacy arising from no shelter.²²²

However, the German court's judgment is a prime example of providing government assistance and safeguarding the right to an adequate standard of living, instead of targeting the underlying causes of poverty. This reveals the prevalence of redistributive measures as opposed to addressing pre-distribution conditions. Encouraging fruitful discussions about human dignity and the eradication of poverty is a challenge within the current framework. States must prioritise allocating resources to tackle poverty but not feel pressured to forgo their other aims.²²³ Additionally, property protection is accorded axiomatic value.²²⁴ Thus, it is imperative to restructure either the nature of discussion, the nature of property rights or institutional priorities.

The first stage would be to disentangle the meaning of human dignity from working conditions. This suggestion is presented briefly and is not intended to be a comprehensive policy solution. At this stage, this article seeks to focus more on examining the problem instead of providing a solution. This is aimed at catalysing future discussion.

²¹⁹ Scott Leckie and Anne Gallagher (ed), *Economic, Social, and Cultural Rights: A Legal Resource Guide* (University of Pennsylvania Press 2006) xvii.

²²⁰ UN Economic and Social Council (25th Session), 'Statement on Poverty and the ICESCR' (4 May 2001) UN Doc. E/C.12/2001/10 para 9.

²²¹ Kallstrom and Eide (n 57), 495.

²²² Chinkin (n 172), 574-575.

²²³ Peter Uvin, *Human Rights and Development* (Kumarian Press 2004), 184-194.

²²⁴ John Linarelli, Margot E Salomon, and Muthucumaraswamy Sornarajah, *The End of Empire and the Search for Justice – NIEO and Beyond* (Oxford University Press 2018), 99.

There is a significant relevance to Nussbaum's discussion regarding the Capabilities approach and Sen's subsequent contribution thereto.²²⁵ Human capabilities are interpreted as the inquiry about what people *can do and be*, constructing intuitively a life worthy of the dignity of a human being.²²⁶ There is a separation between capabilities, which are political goals that governments ought to promote, and the precise functioning of those capabilities, which entails the choice of citizens on whether and how they ought to pursue those capabilities.²²⁷ A just society should shape public policy in order to secure the actual capabilities, and the ability to exercise them. For example, securing the right to political participation will not only involve giving people the right to vote, but also asking if the people are in a position to exercise this right. When, for example, a woman is violently threatened to not leave her home, even if she is conceptually accorded the right to vote she is practically incapable of exercising such a right.²²⁸ In view of this logic, recognising a right to an adequate standard of living is insufficient if people lack the necessary material and emotional resources to exercise it.

This approach might make us reconsider our perceptions regarding property rights. The nature of property rights changed following the right to work, allowing one to accumulate property in exchange for labour. It changed when slavery was abolished and when human beings were no longer considered property. Therefore, we should consider new ways of consuming, or new ways of organising society, so that everyone is able to meet their needs and live in dignity.

CONCLUSION

Any class system theoretically allows mobility. However, the odds are stacked against it. Children are likely to end up in the same economic quintile as

²²⁵ The capabilities approach was first mentioned in Amartya Sen, *Development as Freedom* (Oxford University Press 1999).

²²⁶ Nussbaum provides an elaborated list of capabilities that is beyond the scope of this paper, but to complete the argument those are their titles: Life, Bodily Health, Bodily integrity, Sense, Imagination and Thought, Emotions, Practical Reason, Affiliation, Other Species, Play, Control Over One's Environment: Martha C. Nussbaum, 'Capabilities as fundamental entitlements: Sen and social justice' (2003) 9(2-3) *Feminist Economics* 33, 40-43; Martha C. Nussbaum, *Women and Human Development: The Capabilities Approach* (Cambridge University Press 2000), 5.

²²⁷ Nussbaum 2000 (n 224), 12.

²²⁸ Nussbaum (n 224), 37-40.

their parents.²²⁹ In terms of substantial methods of poverty eradication, the main responsibility, according to human rights law, is on individuals to pull themselves out of poverty and overcome structural obstacles. This article does not claim that we are in the era of the Poor Laws. Impoverished people today are not placed in a workhouse, but rather, they are on the streets, working indecent jobs, neglected by social security, or trapped in the poverty cycle, despite being conceptually entitled to a holistic human rights framework. Originally conceived of as a right to allow the little man to shape the market, the socialist right to work was interpreted within a capitalist view of individualism as a right to waged work and decent working conditions. In the event of job insecurity, the right to social security does not protect all and is limited in its influence. At the national level, the right to an adequate standard of living, without substantially agreeable methods of implementation, is being interpreted as the right to social security. This creates ambiguity regarding when the state should intervene, and places the primary responsibility on the individual to act and work to alleviate poverty. Most importantly, it centres on the question of whether someone is employed and what their working conditions are, rather than the question of facilitating human dignity by virtue of their humanity.

Therefore, we should consider whether the right to an adequate standard of living, if it even exists, fits in with our overall narrative, or, in the absence of a clear answer, whether the narrative ought to be modified instead.

²²⁹ Rivera (n 3).

Unilateral Extraterritorial Sanctions: The Search for a Jurisdictional Justification under International Law

Anna Sophie Nagel*

ABSTRACT

The US and the EU, amongst other nations, responded to the Russian invasion of Ukraine in February 2022 with a coordinated series of unprecedented sanctions. Expanding upon previous practice, the US sanctions package in particular contained several restrictive measures with extraterritorial application. This article examines the legality of unilateral extraterritorial sanctions imposed by the US, both historically and currently, by reference to the customary international law on state jurisdiction. It is argued that unilateral extraterritorial sanctions exceed the scope of the traditional jurisdictional grounds of territory, nationality, and security. To legitimise the current sanctions regime against Russia, which requires at least limited extraterritoriality to be effective, anti-evasion is proposed as a novel basis of jurisdiction over transactions designed to circumvent otherwise lawful sanctions.

*BA Jurisprudence (Jesus College, University of Oxford) '24. The author would like to extend her gratitude to the members of the editorial board of the present volume. The paper takes into account developments in the current Russia/Ukraine sanctions up until 31 December 2022.

INTRODUCTION

When Russian President Vladimir Putin first moved ‘peacekeeping’ troops into the Donetsk and Luhansk regions, and then launched a full-scale invasion of Ukraine in February 2022, the response of the international community was swift and coordinated. The United States (US) and the European Union (EU), amongst other nations, announced a series of unprecedented sanctions on Russia.¹ The proclaimed aim of these sanctions was to impose costs on Russia in an effort to induce the cessation of its unlawful military activities.² The sanctions packages have since been updated and expanded in response to new developments in the ongoing conflict, notably the purported annexations of four Ukrainian regions in October 2022.

The current sanctions against Russia are a striking illustration of the central importance of unilateral restrictive measures as a global foreign policy tool.³ A key feature of contemporary sanctions practice has been the so-called ‘sanctions design’, which focuses on the art of tailoring sanctions so as to best achieve their objectives.⁴ The adoption of extraterritorial sanctions is one

¹ See generally Chad P. Bown, ‘Russia’s war on Ukraine: A sanctions timeline’ (*Peterson Institute for International Economics*, 27 February 2023) <<https://www.piie.com/blogs/realtime-economics/russias-war-ukraine-sanctions-timeline>> accessed 7 March 2023.

² Joseph R. Biden, ‘Remarks by President Biden on Russia’s Unprovoked and Unjustified Attack on Ukraine’ (White House, Washington D.C., 24 February 2022) <<https://www.whitehouse.gov/briefing-room/speeches-remarks/2022/02/24/remarks-by-president-biden-on-russias-unprovoked-and-unjustified-attack-on-ukraine/>> accessed 7 March 2023; Directorate-General for Neighbourhood and Enlargement Negotiations, ‘Press statement by President von der Leyen on Russia’s aggression against Ukraine’ (24 February 2022) <https://neighbourhood-enlargement.ec.europa.eu/news/press-statement-president-von-der-leyen-russias-aggression-against-ukraine-2022-02-24_en> accessed 7 March 2023.

³ See generally Felbermayr and others, ‘The Global Sanctions Data Base’ (2020) 129 *European Economic Review* 103561; updated in Kirikakha and others, ‘The Global Sanctions Data Base: An Update that Includes the Years of the Trump Presidency’ [2021] in Peter A.G. van Bergeijk (ed), *Research Handbook on Economic Sanctions* (Edward Elgar Publishing 2021).

⁴ Charlotte Beaucillon, ‘An Introduction to Unilateral and Extraterritorial Sanctions: Definitions, State of Practice and Contemporary Challenges’ in Charlotte Beaucillon (ed),

particularly controversial method of modern sanctions design. Unilateral extraterritorial sanctions aim to regulate the conduct of persons located in and bearing the nationality of a third state, especially vis-à-vis sanctioned individuals, entities, or states. Such extraterritorial application of unilateral sanctions, which is particularly prevalent in US practice,⁵ faces criticism from affected states,⁶ practitioners,⁷ and academic commentators.⁸

From a public international law perspective, unilateral extraterritorial sanctions may be criticised as contradicting the customary law on state jurisdiction.⁹ To date, states – especially the US – have primarily defended their extraterritorial sanctions with reference to the established bases of jurisdiction.¹⁰ The present article will examine these jurisdictional justifications for unilateral extraterritorial sanctions with regard especially to the current sanctions against Russia in response to the invasion of Ukraine. It will be argued that the traditional grounds cannot confer jurisdictional validity on unilateral extraterritorial sanctions without an impermissible extension of the underlying principles. To guarantee the international legality of secondary sanctions, a new base of jurisdiction will be submitted: anti-evasion. Anti-evasion should be recognised as a new ground of jurisdictional validity for a narrow category of secondary sanctions only, in order to increase the effectiveness of sanctions as instruments promoting the values of the international community.

Research Handbook on Unilateral and Extraterritorial Sanctions (Edward Elgar Publishing 2021) 4.

⁵ Iryna Bogdanova, 'Unilateral Sanctions in International Law and the Enforcement of Human Rights' (2022) 9 *World Trade Institute Advanced Studies* 89.

⁶ Such as the EU: cf Stoll and others, 'Extraterritorial sanctions on trade and investments and European responses' (2020) European Parliament Think Tank, <[https://www.europarl.europa.eu/RegData/etudes/STUD/2020/653618/EXPO_STU\(2020\)653618_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2020/653618/EXPO_STU(2020)653618_EN.pdf)> accessed 7 March 2023.

⁷ Rathbone, Jeydel and Lentz, 'Sanctions, Sanctions Everywhere: Forging a Path through Complex Transnational Sanctions Laws' (2013) 44 *Georgetown Journal of International Law* 1055.

⁸ Susan Emmenegger, 'Extraterritorial Economic Sanctions and their Foundation in International Law' (2016) 33 *Arizona Journal of International and Comparative Law* 631.

⁹ See generally Charlotte Beaucillon, 'Practice Makes Perfect, Eventually? Unilateral State Sanctions and the Extraterritorial Effects of National Legislation' in Natalino Ronzitti (ed), *Coercive Diplomacy, Sanctions and International Law* (Brill Nijhoff 2016).

¹⁰ Yann Kerbrat, 'Unilateral/Extraterritorial Sanctions as a Challenge to the Theory of Jurisdiction' in Beaucillon (n 4).

In order to unpack the jurisdictional validity of unilateral extraterritorial sanctions, the next section will outline the terminology of key concepts used before an overview of the main extraterritorial sanctions contained in the current US sanctions regimes against Russia in section three. Section IV will assess these extraterritorial sanctions with reference to the traditional bases of jurisdiction in international law: territory, nationality, and security. This will be followed by a conceptual argument to recognise new grounds of jurisdiction in light of the purposes underlying the doctrine of jurisdiction. Sections VI and VII develop the proposal of anti-evasion as a novel ground of jurisdiction justifying extraterritorial sanctions. Section VIII concludes.

II. DEFINITIONAL DELINEATION

In public international law, unilateral or autonomous sanctions are restrictive measures imposed by an individual state (or regional international organisations such as the EU, which for the purposes of this article, will be conceived of as a ‘state’) against another state outside the framework of Chapter VII of the United Nations (UN) Charter (i.e., without or beyond an existing sanctions regime pursuant to a resolution of the UN Security Council).¹¹ Contemporary sanctions packages encompass both primary and secondary sanctions. Primary sanctions impose restrictions or conditions on the targets’ ability to freely conduct economic relations and undertake other (non-economic) activities.¹² Secondary sanctions are designed to enhance the effectiveness of primary sanctions by imposing restrictions intended to deter third parties from entering into, or maintaining relations with, the targets of the primary sanctions.¹³

The concept of ‘jurisdiction’ denotes a state’s competence to regulate the conduct of natural and legal persons¹⁴ and encompasses the legislative, judicial, and executive powers of the state (prescriptive, adjudicative, and

¹¹ Alain Pellet and Alina Miron, ‘Sanctions’, *Max Planck Encyclopedia of Public International Law (MPEPIL)* (2013) para 62.

¹² cf Tom Ruys and Cedric Ryngaert, ‘Secondary Sanctions: A Weapon out of Control? The International Legality of, and European Responses to, US Secondary Sanctions’ [2020] *British Yearbook of International Law* 22.

¹³ Rathbone, Jeydel and Lentz (n 7) 1070.

¹⁴ James Crawford, *Brownlie’s Principles of Public International Law* (9th edn, OUP 2019) 440.

enforcement jurisdiction).¹⁵ In the absence of some specific basis in international law, the starting point is the presumption that jurisdiction may not be exercised without a territorial or personal nexus to the state.¹⁶ Primary sanctions may therefore be extraterritorial to the extent that the ‘designated persons’ whose conduct they purport to regulate are non-nationals located outside of the sanctioning state’s territory. In contrast, secondary sanctions are extraterritorial by definition because they purport to govern the activities of third parties that possess no territorial or personal connection to the sanctioning state.¹⁷

The dominant view in the current academic debate conceptualises both primary and secondary sanctions as an exercise of the jurisdiction to prescribe.¹⁸ ‘Prescriptive jurisdiction’ refers to the authority of the legislative organs of the state to enact binding law; whereas ‘enforcement jurisdiction’ refers to the authority of the executive and juridical organs of the state to enforce such law.¹⁹ The deterrent effect of sanctions enforces the extraterritorially-prescribed norm in a wider sense, but not under the traditional definition of enforcement jurisdiction.²⁰ In any case, the validity of an exercise of enforcement jurisdiction turns in part on the existence of valid prescriptive jurisdiction to regulate the underlying conduct.²¹ This article will therefore analyse the jurisdictional validity of unilateral extraterritorial sanctions as exercises of the prescriptive jurisdiction of the sanctioning state.

¹⁵ Bernard H Oxman, ‘Jurisdiction of States’, *MPEPIL* (2007) para 3.

¹⁶ Cedric Ryngaert, *Jurisdiction in International Law* (2nd edn, OUP 2015) 30. The broad prohibitive rule approach per the PCIJ in SS ‘Lotus’ (France v Turkey) (1927) PCIJ Series A, No 10, 19 has been rejected in state practice and therefore does not form part of customary law (and, in any case, has arguably been rejected by the ICJ in Arrest Warrant of 11 April 2000 (Democratic Republic of the Congo v Belgium), Judgment, ICJ Reports (2002) 3).

¹⁷ Beaucillon (n 4) 6.

¹⁸ See for example Jeffrey A. Meyer, ‘Second Thoughts on Secondary Sanctions’ (2009) 30 University of Pennsylvania Journal of International Law 905, 952-955, accepted in Julia Schmidt, ‘The Legality of Unilateral Extra-territorial Sanctions under International Law’ (2022) 27 Journal of Conflict & Security Law 53, 67.

¹⁹ Malcolm N. Shaw, *International Law* (9th edn, CUP 2021) 559.

²⁰ On the relationship between the jurisdiction to enforce and the jurisdiction to prescribe more generally, see Ryngaert (n 16) 9-10.

²¹ Meyer (n 18) 952.

III. THE EXTRATERRITORIAL REACH OF CURRENT US UKRAINE-RELATED SANCTIONS AGAINST RUSSIA

The current Ukraine-related sanctions against Russia imposed by the US include both primary and secondary extraterritorial sanctions. The legal basis for US sanctions in response to the Russian invasion of Ukraine is a series of Executive Orders (EOs 14065,²² 14066,²³ 14068,²⁴ 14071²⁵) made pursuant to the International Emergency Economic Powers Act (IEEPA)²⁶ and the National Emergencies Act (NEA).²⁷ The EOs vest authority in the Secretary of the Treasury and the Secretary of State to employ all powers granted to the President by the IEEPA.²⁸ The majority of the restrictive measures currently adopted concern 'US persons' and property 'within the US'.²⁹ These may be extraterritorial by virtue of the broad territorial and personal nexuses invoked in US sanctions practice, but not merely because they are intended to burden and deter certain extraterritorial conduct.³⁰

Some US Ukraine/Russia sanctions are extraterritorial by virtue of their design. EOs 14068 and 14071 prohibit the re-export of, respectively, luxury (and other) goods and USD-denominated banknotes³¹ and services³² to persons located in Russia. Such re-export prohibitions necessarily purport to regulate the conduct of foreign-based entities by requiring them to refrain from engaging in the prohibited exports from the territory of a third state.³³ Other

²² 87 Fed. Reg. 10293 (21 February 2022).

²³ 87 Fed. Reg. 13625 (8 March 2022).

²⁴ 87 Fed. Reg. 14381 (11 March 2022).

²⁵ 87 Fed. Reg. 20999 (6 April 2022).

²⁶ International Emergency Economic Powers Act, 91 Stat. 1625, 50 U.S.C. § 1701 et seq (1977) (US).

²⁷ National Emergencies Act, 90 Stat. 1255, 50 U.S.C. § 1601 et seq (1976) (US).

²⁸ Blocking Property of Certain Persons and Prohibiting Certain Transactions With Respect to Continued Russian Efforts To Undermine the Sovereignty and Territorial Integrity of Ukraine Order 2022, EO 14065 (n 22) § 10.

²⁹ As generally authorised by International Emergency Economic Powers Act, 50 U.S.C. § 1702 (2011) (US). cf EO 14065 (n 22) §§ 1(a) and 2(a); EO 14066 (n 23) §§ 1(a)(ii) and (iii); EO 14068 (n 24) §§ 1(a)(ii)-(v); EO 14071 (n 25) § 1(a).

³⁰ Meyer (n 18) 958.

³¹ EO 14068 (n 24) § 1(a)(ii) and (iv).

³² EO 14071 (n 25) § 1(a)(ii).

³³ Bogdanova (n 5) 104-105. See also Richard Gordon, Michael Smyth and Tom Cornell, *Sanctions Law* (Bloomsbury Publishing Plc 2019) 114.

export controls imposed by the Department of Commerce regulate US exports and exports from third countries that use US technology inputs such as equipment, software and blueprints.³⁴ These controls, known as foreign direct product rules (FDPR), are extraterritorial in their application to non-US exports.

The US EOs have also imposed extraterritorial secondary sanctions. For example, the asset freezes under EO 14065 extend to ‘any person’ (i.e. not limited to ‘US persons’) deemed to have ‘materially assisted, sponsored, or provided ... support for ... any person whose property and interests in property are blocked pursuant to this order’.³⁵ In light of the extremely expansive interpretations of similar provisions in previous sanctions regimes adopted by the Office of Foreign Assets Control (OFAC),³⁶ these sanctions will likely have a considerable extraterritorial reach. The US has previously threatened to expand the net of secondary sanctions within the Ukraine/Russia sanctions regime to ensure Russia’s financial isolation.³⁷

Similarly, all EOs prohibit any transaction that ‘evades or avoids, has the purpose of evading or avoiding, causes a violation of, or attempts to violate’ any of the prohibitions, and prohibit conspiracies to violate any of the prohibitions.³⁸ These anti-evasion clauses partially reflect the definition of an ‘unlawful act’ under the IEEPA, which includes conspiring to violate or causing a violation of a sanctions measure.³⁹ The extraterritorial reach of US sanctions is thus achieved by prohibiting any conduct by non-US entities outside of the territory of the US which causes a violation in the US.⁴⁰

³⁴ Martin Chorzempa, ‘New technology restrictions against Russia could also target China’ (*PIIE*, 7 March 2022)

<<https://www.pie.com/blogs/realtime-economic-issues-watch/new-technology-restrictions-against-russia-could-also-target>> accessed 7 March 2023.

³⁵ EO 14065 (n 22) § 2(a)(iv).

³⁶ See for example the prohibition of ‘facilitation’: Rathbone, Jeydel and Lentz (n 7) 1102.

³⁷ James Politi, ‘US threatens to punish third parties helping Moscow evade sanctions’ (*Financial Times*, 25 March 2022).

³⁸ See for example EO 14065 (n 22) § 4(a)-(b).

³⁹ IEEPA (n 26) § 206.

⁴⁰ See generally Gordon, Smyth and Cornell (n 33) 114.

IV. EXTRATERRITORIAL SANCTIONS AND THE TRADITIONAL GROUNDS OF JURISDICTION

4.1 The Territoriality Principle

The territoriality principle provides every state with the right to regulate activities on its territory.⁴¹ Subjective territoriality creates jurisdiction over *conduct* taking place within the state's territory, whereas objective territoriality governs *consequences* of conduct which took place outside its territory.⁴² Combined, the two principles permit the invocation of territorial jurisdiction whenever one of the constituent elements of a prohibited act takes place on the territory of the state claiming jurisdiction.⁴³

The US has relied on different variations of the territoriality principle to justify the jurisdictional validity of its unilateral extraterritorial sanctions. The fines imposed by OFAC against the European banks BNP Paribas (BNPP)⁴⁴ and Crédit Agricole⁴⁵ for violating US sanctions provide an illustration of the US conception of a 'constituent element' for the purposes of territorial jurisdiction in the context of financial transactions. The US established the applicability of its sanctions law to the prohibited transaction (the financing of a transaction with a sanctions target) by BNPP, a non-US financial institution, by reference to, *inter alia*, its jurisdiction over the clearing transaction performed by a financial institution on US territory.⁴⁶ 'Clearing' refers to the preparatory steps undertaken before settlement (i.e., actual performance) of a contracted-for financial transaction, most often by a third-party clearinghouse.⁴⁷ To establish

⁴¹ Shaw (n 19) 561.

⁴² Shaw (n 19) 563.

⁴³ Crawford (n 14) 442.

⁴⁴ See generally *US v BNP Paribas SA* (2014) 14 Cr 460 (SDNY), Statement of Facts and Department of the Treasury, Settlement Agreement between OFAC and BNPP (2014) Doc No COMPL-2013-193659 161, 21-26; as analysed in Bismuth: 'Pour une appréhension nuancée de l'extraterritorialité du droit américain – Quelques réflexions autour des procédures et sanctions visant Alstom et BNP Paribas' (2015) *Annuaire français de droit international* LXI (CNRS Éditions) 785, 788-789, 796.

⁴⁵ See generally Department of the Treasury, Settlement Agreement between OFAC and CA-CIB (2014) Doc No COMPL-1000368.

⁴⁶ *US v BNPP* (n 44).

⁴⁷ cf ECB, 'Glossary of Terms Related to Payment, Clearing and Settlement Systems' (ECB, December 2009)

territorial jurisdiction, the clearing transaction was conceived of as a 'constituent element' of the prohibited transaction rather than as a distinct act. Such territorial 'correspondent account-based jurisdiction'⁴⁸ has been upheld by US courts.⁴⁹

It is likely that this justification will be employed to the current Ukraine-related sanctions against Russia. For example, EO 14068 prohibits 'any approval, financing, facilitation, or guarantee by a United States person, wherever located, of a transaction by a foreign person where the transaction by that foreign person would be prohibited by this section if performed by a United States person or within the United States'.⁵⁰ Because the definition of a 'US person' includes 'any person in the United States',⁵¹ the US could invoke territorial jurisdiction justified by its 'constituent element' reasoning, as above, to fine a non-US entity for financing a prohibited transaction under the Order if the clearing transaction was done by any financial institution on US territory.

It is highly doubtful whether such reasoning confers jurisdictional validity on US extraterritorial sanctions. The constituent element analysis of territorial jurisdiction in public international law derives from the model of criminal law, in which constituents of a crime must be definitionally essential to the commission of the crime.⁵² Thus defined, a clearing transaction performed on US territory does not form a 'constituent element' of the prohibited financing transaction: it is a distinct and incidental act.⁵³ Therefore, grounding jurisdiction on the mere routing of financial messages via US servers without any other link with the US whatsoever cannot be compatible with territorial jurisdiction.⁵⁴

<<https://www.ecb.europa.eu/pub/pdf/other/glossaryrelatedtopaymentclearingandsettlementsystems.pdf>> accessed 9 March 2023, 5.

⁴⁸ Explicitly termed such in 87 Fed. Reg. 38939 (7 August 2018) § 2.

⁴⁹ See for example *US v Budovsky* (2015) SDNY 13 Cr 368 (DLC); *Licci v Lebanese Canadian Bank SAL* (2016) 732 F3d 161 (2nd Cir); and especially *US v Reza Zarrab* (2016) SDNY 15 Cr 867 (RMB).

⁵⁰ EO 14068 (n 24) § 1(a)(v).

⁵¹ *ibid* § 4(d).

⁵² Halsbury's Laws (5th ed, 2018) Vol 61, para 213.

⁵³ See Bismuth (n 44) 796.

⁵⁴ See generally Ruys and Ryngaert (n 12) 22.

It must be noted, however, that such ‘incidental-act’ territorial jurisdiction is not limited to the sanctions context: for example, the Foreign Corrupt Practices Act (FCPA)⁵⁵ applies to all ‘issuers’ which register securities with the US Securities and Exchange Commission (SEC),⁵⁶ creating an ‘exchange-based jurisdictional system’⁵⁷ over *foreign persons’ foreign acts* of bribery based on the mere fact of listing on a US stock exchange. Routing payments through US bank accounts or sending emails to US companies has been found sufficient to establish such jurisdiction.⁵⁸ This practice is in tension with the US Restatement (Third) of Foreign Relations Law which requires the conduct in question to take place ‘wholly or in *substantial part*’ within a state’s territory to establish its prescriptive jurisdiction⁵⁹ (which may be understood as an expression of *opinio juris*).

The ‘effects doctrine’ is another variation of the objective territoriality principle⁶⁰ relied on by the US to justify the extraterritorial application of its unilateral sanctions. Originally developed in US antitrust law,⁶¹ this doctrine creates jurisdiction over extraterritorial acts which cause some harmful effect in the territory of the prescribing state.⁶² One of the most famous invocations of the ‘effects principle’ in US sanctions practice is the Helms-Burton Act,⁶³ Title III of which was activated in May 2019 by US President Trump after 22 years of

⁵⁵ Pub. L. No. 95-213, 91 Stat 1494 (1977) (codified at 15 USC §§ 78a et seq).

⁵⁶ FCPA § 78dd-1(a), in conjunction with §§ 78I and 78o(d).

⁵⁷ Ryngaert (n 16) 87.

⁵⁸ See for example *US v Technip SA (Plea Agreement)* SD Tex, No 4:10-CR-00439 (2010). Technip had bribed Nigerian officials through UK and Japanese agents, routing payments through US banks. It settled with the DOJ and SEC for USD338 million.

⁵⁹ American Law Institute, ‘*Restatement of the Law Third, The Foreign Relations Law of the United States*’ (St Paul 1987) § 402(1)(a); not superseded by *Restatement Fourth* (2018).

⁶⁰ Simma and Müller, ‘Exercise and Limits of Jurisdiction’ in Crawford and Koskeniemi (eds), *The Cambridge Companion to International Law* (CUP 2012) 140; Kerbrat (n 10) 176 et seq. Note that some authors conceptualise the effects doctrine as an independent basis of jurisdiction: for example Crawford (n 14) 447.

⁶¹ ‘It is settled law [...] that any State may impose liabilities, even upon persons not within its allegiance, for conduct outside its borders that has consequences within its borders which the State reprehends; and these liabilities other States will ordinarily recognise.’ in *US v Aluminium Co of America* 148 F2d 416 (1945) (2nd Cir). See generally Basedow, ‘Antitrust or Competition Law, International’, *MPEPIL* (2014).

⁶² Crawford (n 14) 447.

⁶³ Cuban Liberty and Democratic Solidarity (LIBERTAD) Act (‘Helms-Burton Act’) Pub. L. No. 104-114, 110 Stat 785 (1996) (codified at 22 USC §§ 6021 et seq).

suspension.⁶⁴ Section 103 of the Act claims that '[i]nternational law recognizes that a nation has the ability to provide for rules of law with respect to conduct outside its territory that has or is intended to have substantial effect within its territory'. Accordingly, the adverse effects on US commerce and US policy⁶⁵ were the basis for the extraterritorial⁶⁶ application of the Section 302 prohibition on 'trafficking' in US property. The invocation of the 'effects doctrine' to establish the jurisdictional validity of unilateral extraterritorial sanctions has been upheld by US courts.⁶⁷

The US has not explicitly relied on the 'effects doctrine' to justify the jurisdictional validity of its current Ukraine/Russia sanctions regime. However, the EOs prescribing the sanctions describe the national emergency caused by the Russian invasion as 'an unusual and extraordinary threat to the national security and foreign policy of the United States'.⁶⁸ If, as will be explored below, the scope of the protective principle is properly limited to national security and the government functions of the state,⁶⁹ this reference to foreign policy may be interpreted as an implicit invocation of the 'effects doctrine', by partial analogy to the justification for extraterritoriality under the Helms-Burton Act. It is therefore possible that the US will seek to justify its current sanctions against Russia by relying on the doctrine.

⁶⁴ The White House, 'President Donald J Trump is Taking a Stand For Democracy and Human Rights in the Western Hemisphere' (*The White House*, 17 April 2019) <<https://trumpwhitehouse.archives.gov/briefings-statements/president-donald-j-trump-taking-stand-democracy-human-rights-western-hemisphere/>> accessed 8 March 2023.

⁶⁵ In a letter addressed to Senator Kennedy in 1979, the Secretary of State explained that: 'We have justified our actions [in applying an embargo extraterritorially] as needed to prevent adverse effects on US commerce or an evasion of important US policy', reproduced in Nash, 'Contemporary Practice of the United States Relating to International Law' (1980) 74 AJIL 182. See generally IEEP (n 26) § 202 (codified at 50 USC § 1701).

⁶⁶ On the extraterritorial application of the Helms-Burton Act and its justification by reference to the effects doctrine, see Kerbrat (n 10) 176-177.

⁶⁷ See for example *US v Budonsky* (n 49) ('[a] jurisdictional nexus exists "when the aim of that activity is to cause harm inside the United States or to US citizens or interests"').

⁶⁸ See for example the preamble to EO 14066 (n 23).

⁶⁹ 'Protective jurisdiction is proper if the activity threatens the security or government functions of the United States': *US v Peterson* (1987) 812 F2d 486, 494 (9th Cir). See also *Arrest Warrant* (n 16) 92 (Judge Rezek) ('... the principle of the defence of certain legal interests to which the State attaches particular value: the life and physical integrity of the sovereign, the national heritage, good governance').

However, the ‘effects doctrine’ cannot justify the extraterritorial application of current US sanctions against Russia. Although the existence of the doctrine and its applicability to unilateral sanctions is generally accepted in both state practice⁷⁰ and international case law,⁷¹ the limits of the doctrine under public international law are contested but undoubtedly narrower than claimed in US sanctions practice. In particular, effects jurisdiction is triggered only if the regulated conduct has ‘direct, foreseeable and substantial effects’ in the prescribing state’s territory.⁷² It is difficult to see, however, how US-sanctioned economic activities by foreign persons outside of US territory could produce such effects in US territory.⁷³ While foreseeable, any adverse effect could only be indirect, i.e. resulting from continuing support for the Russian Government and its policy of aggression, which the sanctions seek to undermine. Moreover, while theoretically possible, it is highly unlikely that a sanctioned transaction would have a *substantial* indirect effect on *US territory* in the absence of any other link. In ordinary circumstances, a transaction between third parties affects neither the US economy nor the integrity of the US sanctions regime domestically.

Moreover, it is particularly controversial whether effects-based jurisdiction applies when the conduct in question complies with the laws of the state in which it was carried out.⁷⁴ As Emmenegger points out, this point distinguishes the sanctions context from the field of antitrust regulation,⁷⁵ in which the ‘effects doctrine’ was first developed: whereas antitrust law rests on similar shared premises in different states, this is not true for most sanctions regimes, given their political and polarising nature. However, because of the

⁷⁰ Kerbrat (n 10) 177-178. For explicit EU acceptance in the sanctions context, see Stoll and others (n 6) 53.

⁷¹ *France v Turkey* (n 16) 23; *Arrest Warrant* (n 16) 77 (Judges Higgins, Kooijmans, Buergenthal).

⁷² As accepted in competition law: *Hartford Fire Ins Co v California* (1993) 509 US 764, 796 (SCOTUS) and Case C-413/14 *Intel Corporation Inc v European Commission* [2017] ECLI:EU:C:2017:632 paras 40ff; and invoked in relation to US unilateral sanctions in response to the Euro-Siberian pipeline crisis: *Campagne Européenne Des Petroles S.A. v Sensor Nederland BV* (1983) (Hague District Court) in (1983) 22 *International Legal Materials* 66, 72.

⁷³ cf Thilo Rensmann, ‘Völkerrechtliche Grenzen extraterritorialer Wirtschaftssanktionen’ in Dirk Ehlers and Hans-Michael Wolfgang (eds), *Recht der Exportkontrolle: Bestandsaufnahme und Perspektiven* (Deutscher Fachverlag 2015), 106; Emmenegger (n 8) 657 (assessing jurisdiction in the BNP Paribas case).

⁷⁴ Emmenegger (n 8) 656.

⁷⁵ *ibid* 656.

remarkable unity of international responses to Russian aggression, this objection may be less weighty in the particular context of Russia/Ukraine sanctions.

4.2 The Nationality Principle

By virtue of the nationality or ‘active personality’ principle, every state possesses jurisdiction over acts of its (natural or legal) nationals, wherever located.⁷⁶ The application of the principle may be extended through reliance on residence and other connections as evidence of allegiance to the state owed by non-nationals.⁷⁷ Although the conditions for the grant of nationality are prescribed at the domestic level, international law appears to require a ‘genuine connection’ with the state.⁷⁸

The US has relied on several extensive interpretations of the nationality principle in its extraterritorial sanctions practice. Under certain sanctions programmes, compliance is required not only from ‘US persons’ strictly speaking, but also from foreign companies owned or controlled by US companies.⁷⁹ For example, under Section 218 of the Iran Threat Reduction and Syria Human Rights Act,⁸⁰ the civil penalties provided for by the IEEPA shall be imposed on US persons for sanctions violations committed by an entity ‘owned or controlled by the United States person and established or maintained outside the United States’. ‘Owned or controlled’ is broadly interpreted in the same section to mean holding more than 50 percent of the equity interest or a majority of the seats on the board of directors, or otherwise controlling the actions, policies and personnel decisions of the entity. This ‘control’ criterion for personal jurisdiction was reiterated in relation to sanctions reimposed against Iran after US withdrawal from the Joint Comprehensive Plan of Action (JCPOA).⁸¹

⁷⁶ *SS ‘Lotus’* (n 16) 92 (Judge Moore); See generally Crawford (n 14) 443-444.

⁷⁷ Crawford (n 14) 443.

⁷⁸ *Nottebohm Case (Liechtenstein v Guatemala)* (Second Phase), Judgment, ICJ Reports (1955) 4, 23. This requirement is now generally accepted: Oxman (n 15).

⁷⁹ US Department of the Treasury, OFAC Consolidated Frequently Asked Questions, ‘Who must comply with OFAC regulations?’ (*US Department of the Treasury*, 15 January 2015), <<https://home.treasury.gov/policy-issues/financial-sanctions/faqs/11>> accessed 9 March 2023.

⁸⁰ Pub. L. No. 112-158, 126 Stat. 1214 (2012) (codified at 22 USC § 8701 et seq).

⁸¹ EO 13846 (n 48) § 8.

The US Ukraine/Russia sanctions package does not currently extend to foreign subsidiaries and therefore does not require the invocation of the control criterion. A juridical ‘US person’ is defined in the EOs prescribing the sanctions as an ‘entity organized under the laws of the United States or any jurisdiction within the United States (including foreign branches)’.⁸² The Directives 1A and 2 under EO 14024 make explicit that the sanctions apply to ‘US subsidiaries’ only, which includes ‘those branches, offices, and agencies of foreign financial institutions that are located in the United States, but not such institutions’ foreign branches, offices, or agencies’.⁸³ The core personal nexus is maintained because a branch, unlike a subsidiary, is not a separate legal entity but rather an extension of the parent company (and therefore not independently incorporated under a non-US jurisdiction). The sanctions therefore cannot be considered extraterritorial by reference to the nationality of the corporations regulated.

In light of previous US practice, for example concerning Iran,⁸⁴ it is however possible that the Ukraine-related sanctions against Russia will be extended to controlled foreign subsidiaries. Such an extension would not be justified under the international law of jurisdiction. The control criterion for personal jurisdiction is employed exclusively by the US and is therefore not generally accepted in state practice.⁸⁵ The lack of international protest⁸⁶ is likely a consequence of the overlap in the sanctions regimes of the US and other states, which puts non-US subsidiaries in a similar position under both extraterritorial US sanctions law and other sanctions regimes. Moreover, control-based jurisdiction contradicts the method of identification of corporate nationality laid down by the ICJ in *Barcelona Traction* and *Abmadou Sadio Diallo*, which rests on

⁸² See EO 14065 (n 22) § 8(c).

⁸³ OFAC, Directive 1A under Executive Order 14024 (22 February 2022), ‘Prohibitions Related to Certain Sovereign Debt of the Russian Federation’ and Directive 2 under Executive Order 14024 (24 February 2022), ‘Prohibitions Related to Correspondent or Payable-Through Accounts and Processing of Transactions Involving Certain Foreign Financial Institutions’.

⁸⁴ The Iran Threat Reduction and Syria Human Rights Act strengthened already existing sanctions against Iran.

⁸⁵ See generally Kerbrat (n 10) 171-175.

⁸⁶ cf for example the absence of any criticism in Stoll and others (n 6).

the place of incorporation rather than the nationality of the shareholders.⁸⁷ In any case, it is difficult to see how the control criterion satisfies the requirement of a 'genuine connection' between the foreign controlled entity and the sanctioning state. A subsidiary incorporated under, and wholly operating in, a different jurisdiction does not evidence a 'social fact of attachment' to the sanctioning state, and does not possess reciprocal rights and duties in relation to the sanctioning state,⁸⁸ as evidence of such a 'genuine link'.

Implicitly, the US has also relied on the nationality principle to justify its unilateral extraterritorial sanctions consisting of export and re-export prohibitions. Since the 1982 Euro-Siberian pipeline controversy,⁸⁹ the US has asserted that its jurisdiction 'runs with'⁹⁰ exported US goods or technology at all times after these leave US territory. This reasoning has been expanded to cover certain foreign-origin items which incorporate US goods or technology.⁹¹ The US now purports to prescribe to '(a) wholly non-US parties, (b) who are located outside the United States and who possess lawfully obtained US origin items, lawfully produced foreign-origin items with non-*de minimis* US content, or certain wholly non-US origin items, (c) not only where and to whom they may provide these items, but also (d) what may not be done with these items in-country, and (e) to do so in perpetuity'.⁹² This assertion of jurisdiction over re-exports is considered to invoke the nationality principle with respect to exported goods and technology.⁹³

⁸⁷ *Barcelona Traction, Light and Power Company, Limited* (Belgium v Spain) (Judgment) [1970] ICJ Rep 3, 28-47; *Ahmadou Sadio Diallo* (Republic of Guinea v Democratic Republic of the Congo) (Preliminary Objections, Judgment) [2007] ICJ Rep 582 paras 89-94. See generally Ruys and Ryngaert (n 12) 18.

⁸⁸ *Nottebohm* (n 78) 23 ('[N]ationality is a legal bond having as its basis a social fact of attachment, a genuine connection of existence, interests and sentiments, together with the existence of reciprocal rights and duties. It may be said to constitute the juridical expression of the fact that the individual upon whom it is conferred ... is in fact more closely connected with the population of the State conferring nationality than with that of any other State').

⁸⁹ See generally Bowman, 'A Prescription for Curing US Export Controls' (2014) 97 *Marquette Law Review* 599, 630-639.

⁹⁰ Douglas Rosenthal and William M. Knighton, *National Laws and International Commerce: The Problem of Extraterritoriality* (Routledge 1982) 54.

⁹¹ Commerce and Foreign Trade, 15 C.F.R. § 734.3.

⁹² Bowman (n 89) 644.

⁹³ See for example Patrick C.R. Terry, 'Enforcing US Foreign Policy by Imposing Unilateral Secondary Sanctions: Is Might Right in Public International Law?' (2020) 30

The US has imposed extraterritorial export and re-export prohibitions as part of the current Ukraine/Russia sanctions regime and thereby implicitly relied on the nationality principle as the basis for its jurisdiction. For example, EO 14065 prohibits the ‘exportation, re-exportation, sale, or supply, directly or indirectly, from the United States, or by a United States person ... of any goods, services or technology’ to the Donetsk and Luhansk regions of Ukraine.⁹⁴ The US Department of Commerce, acting through the Industry and Security Bureau, established two new foreign direct product rules (FDPR) which restrict exports of foreign-origin products that use US technology to Russia and Russian ‘military end users’ in particular.⁹⁵ Other export and re-export restrictions have been imposed on luxury goods and USD-banknotes.⁹⁶

By extension, both primary and secondary⁹⁷ US sanctions restricting the export of services may also be considered to invoke the nationality principle. In the context of financial transactions, ‘correspondent account-based jurisdiction’ has been further justified by conceptualising the ‘execution on behalf of others of money transfers’ as an ‘exportation of a service’ from the US, and therefore falling within the scope of extraterritorially applicable sanctions law.⁹⁸ The US implicitly invoked this reasoning against BNPP when it relied on a generalised *lex monetae*, i.e. the use of US currency,⁹⁹ as one of the grounds of jurisdiction justifying the applicability of its sanctions law to a European financial institution. The same reasoning might be applied in the current Ukraine/Russia context in relation to EO 14071, which *inter alia* prohibits the ‘exportation, re-exportation, sale, or supply, directly or indirectly, from the United States, or by a United State person ... of any category of

Washington International Law Journal 1, 15; Laurent Hoff, *Transaktionen in US-Dollar und Sekundäre Sanktionen: Ist die US-amerikanische Sanktionspraxis mit dem Völkerrecht vereinbar?* (Nomos 2019) 147.

⁹⁴ EO 14065 (n 22) § 1(a)(iii).

⁹⁵ 15 C.F.R. §§ 734, 738, 740, 742, 744, 772. See Bureau of Industry and Security, Final Rule, Federal Register, Vol 87, No 42 (3 March 2022).

⁹⁶ EO 14068 (n 24) § 1(a)(ii).

⁹⁷ Mark Jaeger, ‘Circumventing Sovereignty: Extraterritorial Sanctions Leveraging the Technologies of the Financial System’ (2021) 27 Swiss Political Science Review 180, 183.

⁹⁸ *US v Banki* (2012) 2d Cir 685 F3d 99; *US v Homa International Trading Corp* (2004) 2d Cir 387 F3d 144 as cited in *US v Reza Zarrah* (n 49) 17.

⁹⁹ *US v BNPP* (n 44) paras 5, 8, 16 et seq; as analysed in Bismuth (n 44).

services as may be determined ... to any person located in the Russian Federation'.¹⁰⁰

The nationality principle cannot create jurisdiction over extraterritorial prohibitions on exports and re-exports of US origin or foreign goods and services. Item origin-based nationality jurisdiction has generally been rejected as a basis for prescriptive jurisdiction.¹⁰¹ The limited exceptions for marine vessels, aircraft and spacecraft¹⁰², and certain cultural property¹⁰³ are best viewed as *sui generis* rules rather than instances of the nationality principle.¹⁰⁴ This conclusion is reinforced by the characterisation of personal jurisdiction as based on citizens' allegiance¹⁰⁵ and 'interests and sentiments'¹⁰⁶ towards the state: it is therefore logically inconsistent to apply the nationality principle to non-sentient exports. Similarly, goods and services cannot possess rights against, and duties towards, the state.¹⁰⁷ In any case, the concept of permanent nationality of exported items is not applied in other areas of international trade law, including by the US: there is no general acceptance of the concept in state practice, regardless of the paucity of strong objections to US export and re-export sanctions.¹⁰⁸

4.3 The Protective Principle

The protective or security principle, as a corollary of state sovereignty and independence,¹⁰⁹ permits the assumption of jurisdiction over extraterritorial acts which affect the internal or external security or other key interests of the state.¹¹⁰ Although the existence of the doctrine is well-established, its boundaries

¹⁰⁰ EO 14071 (n 25) § 1(a)(ii).

¹⁰¹ Bowman (n 89) 654. In the sanctions context, see Ruys and Ryngaert (n 12) 20.

¹⁰² Oxman (n 15); Crawford (n 14) 448-450.

¹⁰³ Francesco Francioni, 'Cultural Heritage' (2020) *MPEPIL* para 2.

¹⁰⁴ See for example the classification in Crawford (n 14) Ch 21.

¹⁰⁵ Crawford (n 14) 443.

¹⁰⁶ *Nottebohm* (n 78).

¹⁰⁷ *ibid.*

¹⁰⁸ See Sarah H. Cleveland, 'Norm Internalization and US Economic Sanctions' (2001) 26 *Yale Journal of International Law* 1, 6-7.

¹⁰⁹ Ryngaert (n 16) 115.

¹¹⁰ Crawford (n 14) 446.

remain unclear¹¹¹ and malleable.¹¹² State practice explicitly invoking protective jurisdiction is restricted to ‘acts that severely jeopardise a state’s government functions’¹¹³ such as treason, though jurisdiction over currency, immigration and economic offences in US and UK practice may implicitly rely on the principle.¹¹⁴

The US has previously relied on the protective principle to justify the imposition of extraterritorial secondary sanctions on nationals of third states. For example, concerning Russia, the Countering America’s Adversaries Through Sanctions Act (CAATSA)¹¹⁵ required the US President to impose sanctions on, *inter alia*, any person who ‘knowingly ... engages in a significant transaction with a person that is part of, or operates for or on behalf of, the defense or intelligence sectors of the Government of the Russian Federation’¹¹⁶ and any person who ‘knowingly ... makes [such] an investment [into], or sells, leases, or provides to the Russian Federation, for the construction of Russian energy export pipelines, [such] goods, services, technology, information, or support’¹¹⁷ as specified. These secondary sanctions have been partially justified as a response to ‘cyber intrusions and attacks’¹¹⁸ by Russia, with President Trump explicitly invoking Russian ‘interference in [US] democratic process’ and its ‘subversion and destabilization’ of the international community.¹¹⁹ US courts have confirmed that the extraterritorial application of the CAATSA sanction is

¹¹¹ Shaw (n 19) 573.

¹¹² Crawford (n 14) 446 (‘The categories of what may be considered a vital interest for the purposes of protective jurisdiction are not closed, and no criteria exist for determining such interests beyond a vague sense of gravity.’)

¹¹³ Simma and Müller (n 60) 144.

¹¹⁴ Crawford (n 14) 446.

¹¹⁵ Pub. L. No. 115-44 (2017) (codified at 22 USC § 9401 et seq); especially the Countering Russian Influence in Europe and Eurasia Act (CRIEEA) (22 USC § 9501 et seq).

¹¹⁶ CAATSA § 231(a) (22 USC § 9525).

¹¹⁷ CAATSA § 232(a) (22 USC § 9526).

¹¹⁸ CAATSA § 212(3) (22 USC § 9502).

¹¹⁹ The White House, ‘Statement by President Donald J Trump on Signing the “Countering America’s Adversaries Through Sanctions Act”’ (*The White House*, 2 August 2017), <<https://trumpwhitehouse.archives.gov/briefings-statements/statement-president-donald-j-trump-signing-countering-americas-adversaries-sanctions-act/>> accessed 9 March 2023.

motivated by national security concerns and is therefore justified under the protective principle of jurisdiction.¹²⁰

The current extraterritorial sanctions against Russia in response to the invasion of Ukraine similarly invoke the protective principle. The EOs prescribing the sanctions maintain that Russia's purported recognition of the Donetsk People's Republic (DNR) and Luhansk People's Republic (LNR) and its war against Ukraine threaten the 'peace, stability, sovereignty, and territorial integrity of Ukraine', and thereby 'constitute an unusual and extraordinary threat to the national security and foreign policy of the United States'.¹²¹ Under the framework of the IEEPA and NEA, this national emergency authorises the President to impose sanctions. Notably however, the press statements accompanying the new sanctions did not reference domestic security, rather emphasising the threat to 'global peace and stability'¹²² caused by the Russian invasion.

The protective principle cannot justify unilateral extraterritorial sanctions beyond a very narrow set of circumstances. The legitimate exercise of protective jurisdiction is limited to conduct which threatens 'security or essential government functions'¹²³ and cannot therefore be based on indirect threats to the prescribing state through challenges to its foreign policy. The doctrine is further limited to the protection of *national* (i.e. domestic) rather than *international* security.¹²⁴ Thus, only the cybersecurity sanctions against Russia under CAATSA are jurisdictionally valid under the protective principle because they are

¹²⁰ See for example *US v Reza Zarrah* (n 49) in relation to CAATSA sanctions against Iran; as analysed in Susan Emmenegger and Thirza Döbeli, 'The Extraterritorial Application of US Sanctions Law' in Andrea Bonomi and Krista K. Schefer (eds), *US Litigation Today: Still a Threat For European Businesses or Just a Paper Tiger?* (Schulthess Éditions Romandes 2018).

¹²¹ cf preamble to EO 14065 (n 22) and preamble to EO 14066 (n 23).

¹²² See for example The White House, 'Remarks by President Biden Announcing US Ban on Imports of Russian Oil, Liquefied Natural Gas, and Coal' (*The White House*, 8 March 2022) <<https://www.whitehouse.gov/briefing-room/speeches-remarks/2022/03/08/remarks-by-president-biden-announcing-u-s-ban-on-imports-of-russian-oil-liquefied-natural-gas-and-coal/>> accessed 9 March 2023.

¹²³ Meyer (n 18) 938.

¹²⁴ Shaw (n 19) 573 ('The [protective] principle provides that states may exercise jurisdiction over aliens who have committed an act abroad which is deemed prejudicial to the security of the particular state concerned').

responses to electoral interference, which threatens an ‘essential government function’. The invocation of threats to the peace and security of the international community cannot create protective jurisdiction over other CAATSA sanctions or the current sanctions in response to the invasion of the Ukraine.¹²⁵ Moreover, it is difficult to see how a localised armed conflict on the territory of a foreign state, as the war in Ukraine may currently be described, can constitute a threat to the national security of the US as a non-party to the conflict.

V. TRADITIONAL AND NEW GROUNDS OF JURISDICTION: A CONCEPTUAL FRAMEWORK

The discussion above shows the relative usefulness¹²⁶ of the traditional bases of jurisdiction in assessing the jurisdictional validity of unilateral extraterritorial sanctions: although it has been concluded that, on balance, the traditional grounds cannot justify extraterritorial sanctions as implemented by the US, uncertainty remains about the precise limits of the doctrines invoked, particularly in light of conflicting state practice. Similarly, scholars have put forward conflicting views about the extent to which US sanctions (prior to the Russo-Ukrainian conflict) were justified under the traditional principles of jurisdiction. Ruys and Ryngaert conclude that sanctions restricting access to the sanctioning state’s economic and financial systems are justified by the territoriality principle, whereas measures imposing penalties are more problematic.¹²⁷ Meyer argues, more expansively, that a combined ‘territorial’ ground of jurisdiction justifies restrictions on economic relations with third-country nationals which do business with the sanctions target.¹²⁸ Emmenegger contends that most extraterritorial sanctions are justified by the traditional grounds of jurisdiction, but criticises ‘correspondent account-based jurisdiction’ as extending beyond such grounds.¹²⁹

Especially in light of such uncertainty, the delineation of the appropriate scope of jurisdiction must be informed by the underlying purpose of the doctrine of jurisdiction under international law. Jurisdiction, as noted

¹²⁵ As explicitly recognised by the EU: Stoll and others (n 6) 53-54.

¹²⁶ cf Kerbrat (n 10) 168.

¹²⁷ Ruys and Ryngaert (n 12) 9-29.

¹²⁸ Meyer (n 18).

¹²⁹ Emmenegger (n 8).

above, denotes state regulatory competence. By delineating the respective (often conflicting) spheres of competence of individual states, the principles of jurisdiction serve the dual purpose of facilitating effective international governance while simultaneously protecting and giving expression to state sovereignty (and its corollary of sovereign equality).¹³⁰ Sovereignty may be defined as the power and authority states have at any given moment in the development of the international legal system.¹³¹ To exercise sovereignty, a state must be a subject of international law. Under Article 1 of the Montevideo Convention on Rights and Duties of States 1933,¹³² a state as a person (and therefore subject) of international law should possess the following: a) a permanent population, b) a defined territory, c) a government, and d) the capacity to enter into relations with other states. Accordingly, the traditional bases of jurisdiction examined above may be conceptualised as instantiations of the Montevideo criteria of statehood: the territoriality principle reflecting territory, the nationality principle reflecting population, and the protective principle reflecting the existence of a government and its capacity to effectively enter into relations with other states.

A focus on the dual purpose of jurisdiction explains the evolutive nature of the concept of jurisdiction, and suggests that new grounds of jurisdiction may be legitimately recognised. The principle of sovereignty is not absolute. Its content and implications shift in accordance with historical developments, which condition the function of sovereignty in the international legal order at any given time and space.¹³³ It is therefore possible to recognise new grounds of jurisdiction which are not reflective of the core elements of statehood (and indeed *prima facie* restrict sovereignty) if such grounds increase the effectiveness of international governance. For example, the development of the universality principle as conferring jurisdiction over the prosecution of certain international crimes (such as crimes against humanity and war crimes),¹³⁴ arguably served the governance purpose of strengthening the enforcement of

¹³⁰ cf Ryngaert (n 16) 5.

¹³¹ cf Samantha Besson, 'Sovereignty', (2011) *MPEPIL* paras 1, 56-58.

¹³² Which is generally accepted as the most authoritative formulation of the criteria of statehood: Shaw (n 19) 182. cf Opinion No 1 (1991) Yugoslavia Peace Conference Arbitration Commission 1991.

¹³³ Besson (n 131) paras 5, 8-9, 59-84.

¹³⁴ Oxman (n 15).

international human rights law and thereby protecting the rights of individuals. This purpose justified a potential encroachment on the concept of sovereignty.

VI. ANTI-EVASION AS NOVEL JURISDICTIONAL GROUND

A potential new development in the principles of jurisdiction under international law is the recognition of anti-evasion as a jurisdictional ground. This trigger has been recently invoked in the extraterritorial application of EU legislation and upheld judicially in conjunction with other traditional bases of jurisdiction.¹³⁵ The doctrine is ‘intended to catch artificial behaviour designed to evade obligations laid down in EU law’¹³⁶ by requiring third-country entities to comply with EU law where this is ‘necessary or appropriate to prevent the evasion of any provision’¹³⁷ in the relevant regulations.

The US has not sought to justify its imposition of extraterritorial secondary sanctions by reference to anti-evasion as a basis of jurisdiction. Nevertheless, the prescriptive frameworks under which sanctions are currently imposed clearly pursue anti-evasion as a rationale, in order to achieve a broad and therefore effective¹³⁸ application of the respective sanctions regimes. For example, section 228 of CAATSA explicitly provides for sanctions against ‘foreign sanctions evaders’ by amending the Support for the Sovereignty, Integrity, Democracy and Economic Stability of Ukraine Act (SSIDESAs)¹³⁹ to sanction any foreign person who ‘facilitates a significant transaction or transactions, *including deceptive or structured transactions*’¹⁴⁰ for or on behalf of Russian sanctions targets.

¹³⁵ *L’Oréal S.A. and Others v eBay International AG and Others* (2011) Case C-324/09 (CJEU) as interpreted in Lena Hornkohl, ‘The Extraterritorial Application of Statutes and Regulations in EU Law’ (2022) 1 MPILux Research Paper Series 1, 22-23.

¹³⁶ Joanne Scott, ‘The New EU “Extraterritoriality”’ (2014) 51 Common Market Law Review 1343, 1359. See also Hornkohl (n 135).

¹³⁷ See for example Regulation (EU) No 648/2012 of the European Parliament and of the Council on OTC derivatives, central counterparties and trade repositories (4 July 2012), Art. 11(12); Regulation (EU) No 600/2014 of the European Parliament and of the Council on markets in financial instruments (15 May 2014), Art. 28(2).

¹³⁸ Beaucillon (n 4) 5.

¹³⁹ Pub. L. No. 113-95, codified at 22 USC § 8901 et seq (2014).

¹⁴⁰ SSIDESAs § 10(a)(2) (as amended by CAATSA § 228) (emphasis added).

As highlighted above, the current sanctions regime against Russia in response to the invasion of Ukraine also contains explicit anti-evasion clauses. The relevant EOs prohibit, *inter alia*, any transaction that ‘evades or avoids’ or ‘has the purpose of evading or avoiding’¹⁴¹ any of the sanctions prescribed by the Orders. OFAC has stated that it is prepared to use this anti-evasion provision in EO 14024, codified in the Russian Harmful Foreign Activities Sanctions Regulations (RuHSR),¹⁴² to impose sanctions on ‘supporters of Russian sanctions evasion’, such as non-US financial institutions contracting with NSPK, the operator of Russia’s MIR National Payment System.¹⁴³

As a free-standing basis of jurisdiction, anti-evasion may confer jurisdictional validity on a narrow set of unilateral extraterritorial sanctions implemented to prevent the circumvention of the sanctions regime. The rationale of anti-evasion is not contested by other states. In particular, the EU Guidelines on the Implementation and Evaluation of Restrictive Measures (Sanctions) prohibit an EU-incorporated entity from using a company that it controls ‘as a tool to circumvent a prohibition’ and the giving of instructions to the controlled foreign company to such effect.¹⁴⁴ The recognition of anti-evasion as a legal basis for jurisdiction would represent only a limited extension of the current law, since it would apply only to arrangements *designed* to evade sanctions, i.e. deliberately set up with the purpose of circumvention (whether or not in addition to a commercial purpose). In light of this restrictive definition, sanctions targeting ‘innocent’ commercial transactions with a legitimate counterparty would not fall within the scope of anti-evasion jurisdiction.

This limited extension is justified as a tool to increase the effectiveness of unilateral sanctions in promoting the values and interests of the international community, such as the fight against wars of aggression, human rights violations

¹⁴¹ See for example EO 14065 (n 22) § 4.

¹⁴² 31 C.F.R Part 587.

¹⁴³ US Department of the Treasury, OFAC Consolidated Frequently Asked Questions, ‘Do non-US financial institutions risk exposure to sanctions for contracting or otherwise dealing with NSPK?’ (*US Department of the Treasury*, 15 September 2022) <<https://home.treasury.gov/policy-issues/financial-sanctions/faqs/1082>> accessed 9 March 2023.

¹⁴⁴ Council of the EU, ‘Guidelines on Implementation and Evaluation of Restrictive Measures (Sanctions) in the Framework of the EU Common Foreign and Security Policy’ (4 May 2018) Annex to Doc No 5664/18, para 54.

or cyberattacks, without imposing a disproportionate burden on innocent third parties. The need for such an extension of jurisdiction for the purpose of increasing effectiveness of sanctions has already been recognised. For example, in rejecting the effects doctrine as a valid jurisdictional ground in the BNP Paribas case, Emmenegger noted that '[w]hat is affected is the sanctions' global effectiveness, because firms outside the US can engage in the conduct prohibited by US law'.¹⁴⁵ Perhaps paradoxically, the recognition of a narrow anti-evasion principle may decrease the extraterritorial extension of sanctions regimes by preventing gaps and thus generating a comprehensive response to the initial wrongdoing the sanctions seek to deter. In the context of Ukraine-related sanctions against Russia, evasion has been identified as an obstacle to their effectiveness as a method for inducing the Kremlin to fully reverse its actions and end aggression in Ukraine.¹⁴⁶ It is therefore imperative to legitimate sanctioning states' anti-evasion efforts in order to ensure the successful operation of the current sanctions regime. The recognition of anti-evasion as a basis for extraterritorial jurisdiction provides such legitimisation in the context of the international law on state jurisdiction.

Ruys and Ryngaert, in the only existing examination of anti-evasion as a jurisdictional ground in the sanctions context (to the present author's knowledge), conclude that anti-evasion does not jurisdictionally justify the imposition of (secondary) extraterritorial sanctions.¹⁴⁷ Three objections may be discerned from the analysis, all of which can be met. First, it is contended that the anti-evasion jurisdictional ground is specific to the field of financial regulation. However, there is nothing intrinsic in the anti-evasion principle which would prevent its recognition in the sanctions context, independently of developments in financial regulation (and competition law more generally).

¹⁴⁵ Emmenegger (n 8) 657.

¹⁴⁶ International Institute for Strategic Studies, 'Russia and Sanctions Evasion' (2022) 28, Strategic Comment 15
<<https://www.iiss.org/publications/strategic-comments/2022/russia-and-sanctions-evasion>> accessed 9 March 2023. See also Åslund and Snegovaya, 'The Impact of Western Sanctions on Russia and How They Can Be Made Even More Effective' (*Atlantic Council*, 2021),
<<https://www.atlanticcouncil.org/wp-content/uploads/2021/05/The-impact-of-Western-sanctions-on-Russia-and-how-they-can-be-made-even-more-effective-5.2.pdf>>
accessed 9 March 2023.

¹⁴⁷ Ruys and Ryngaert (n 12) 27-28.

Second, it is objected that the principle is too restrictive to be of any value, arguably capturing only transactions with deliberately set up artificial legal vehicles. It is correct that anti-evasion jurisdiction as currently recognised in EU law is limited to *artificial* arrangements *designed* to evade legal obligations, i.e. arrangements which ‘intrinsically lack business rationale, commercial substance or relevant economic justification’¹⁴⁸ beyond this purpose. However, the EU employs the same understanding of jurisdictions in relation to its sanctions,¹⁴⁹ without considering this to be an extension of the current international law on jurisdiction.¹⁵⁰ Accordingly, anti-evasion as a *novel* ground of jurisdiction may be defined less restrictively, thereby giving it practical effect. For example, this author has suggested a definition which captures all transactions *designed* to evade sanctions, whether or not these transactions are deliberately artificial or otherwise lack commercial viability.

Third, the comment that the recognition of anti-evasion as a jurisdictional ground ‘would upgrade the *political rationale* for the imposition of secondary sanctions ... to an independent *jurisdictional ground*’¹⁵¹ may be understood as a principled objection. However, as examined above, the recognition of new jurisdictional grounds for the purpose of advancing political rationales of international governance is in accord with the underlying purposes of the doctrine of jurisdiction. It is therefore not illegitimate to recognise anti-evasion as a jurisdictional ground for the purpose of promoting the effectiveness of sanctions as tools of global governance and measures of enforcement of international law.

VII. RECOGNISING ANTI-EVASION AS A JURISDICTIONAL GROUND

So far, this article has examined the jurisdictional validity of unilateral extraterritorial sanctions by reference to historical and current US practice. It has been argued that US sanctions regimes, including the sanctions against Russia in response to its unlawful aggression in Ukraine, stretch the traditional

¹⁴⁸ Commission Delegated Regulation 285/2014 (13 February 2014) supplementing Regulation (EU) No 648/2012, Art. 3(2).

¹⁴⁹ cf Council of the EU (n 144).

¹⁵⁰ *ibid* para 52 (‘The EU will refrain from adopting legislative instruments having extra-territorial application in breach of international law’).

¹⁵¹ Ruys and Ryngaert (n 12) 27.

bases of jurisdiction beyond their proper limits under customary international law. Thus, the territoriality principle cannot justify so-called ‘correspondent account-based jurisdiction’, nor can it confer, under the ‘effects doctrine’, jurisdictional validity on sanctions against foreign persons who purportedly cause harm to the US within its territory. Personal jurisdiction pursuant to the nationality principle cannot be extended to controlled foreign companies and similarly cannot provide a jurisdictional basis applicable to export and re-export prohibitions against third-country nationals and entities. As a result, the characterisation of the US financial and monetary system as an ‘exported service’ alternatively justifying ‘correspondent account-based jurisdiction’ must be deemed impermissible. Finally, most extraterritorial sanctions imposed by the US do not fall within the proper scope of protective jurisdiction. In particular, unilateral extraterritorial sanctions responding to threats to international peace and security exceed the prescriptive competence of individual states under the protective principle. In short, contrary to US assertions, its model of unilateral extraterritorial sanctions does not conform to the international law of state jurisdiction.

What is the significance of this conclusion? *Prima facie*, the continued prescription (but especially the enforcement) of unilateral extraterritorial sanctions contrary to the principles of jurisdiction amounts to the commission of an international wrong by the US,¹⁵² particularly in light of the considerable adverse impact of US extraterritorial sanctions on third states¹⁵³ (which satisfies the additional requirement of an actualized injury).¹⁵⁴ The characterisation of the extra-jurisdictional sanctions against third parties as countermeasures does not preclude wrongfulness because lawful countermeasures may only be directed against the state which has committed the prior wrong.¹⁵⁵ States affected by US extraterritorial sanctions may therefore invoke the international responsibility of the US on behalf of their targeted nationals. This illegality of US unilateral extraterritorial sanctions under international law, and the consequent possibility

¹⁵² Crawford (n 14) 461.

¹⁵³ See generally Emmanuel Breen, ‘Corporations and US Economic Sanctions: The Dangers of Overcompliance’ in Beaucillon (n 4) 256.

¹⁵⁴ Kimberly Trapp, ‘Jurisdiction and State Responsibility’ in Stephen Allen and others (eds), *The Oxford Handbook of Jurisdiction in International Law* (OUP 2019) 356, 360.

¹⁵⁵ International Law Commission, ‘Draft Articles on the Responsibility of States for Internationally Wrongful Acts’ (2001) Yearbook of the International Law Commission Vol II Part Two, Art. 49(1)-(2).

of an invocation of responsibility, is particularly problematic in the context of Russian aggression against Ukraine because the success of the coordinated action in response at least partially depends on the perceived legitimacy of the measures taken. Otherwise, the political strength of the current sanctions against Russia is vulnerable to an accusation of hypocrisy by Russia and its allies.¹⁵⁶

Accordingly, this article has examined anti-evasion as an alternative novel basis of jurisdiction justifying the imposition of unilateral extraterritorial sanctions. It has been argued that anti-evasion is a viable jurisdictional principle when confined to sanctions explicitly adopted for the purpose of preventing circumvention of the sanctions regime to the state. The limited extension of the current international law of jurisdiction that the recognition of this principle would demand is justified as a way of increasing the effectiveness of sanctions in promoting the values and interests of the international community. In an effort to limit the impact on third parties, the anti-evasion ground of jurisdiction should be confined to conduct designed to circumvent otherwise lawful sanctions.

In practice, the recognition of anti-evasion as a basis of prescriptive jurisdiction in the sanctions context must result from an evolution of the customary international law on jurisdiction. At present, the lack of strong protest by other states against the extraterritoriality of the Ukraine/Russia sanctions imposed by the US is conducive to such a development¹⁵⁷ (though it is insufficient in and of itself to conclusively establish a new customary rule of international law). However, as emphasised by the ICJ in *Nicaragua*, '[i]f a State acts in a way *prima facie* incompatible with a recognised rule, but defends its conduct by appealing to exceptions or justifications contained within the rule itself, then whether or not the State's conduct is in fact justifiable on that basis, the significance of that attitude is to confirm rather than weaken the rule'.¹⁵⁸ In order to implement anti-evasion as a ground of jurisdiction, the US and other states must therefore explicitly rely on it to justify their unilateral extraterritorial

¹⁵⁶ On symbolic significance, see Meyer (n 18) 934-935.

¹⁵⁷ cf for example responses to US sanctions against Iran: Ruys and Ryngaert (n 12) 81-116; Nathaniel Tilahun, 'Resisting (US) Sanctions: A Comparison of Special Purpose Vehicles, Blocking Statutes and Countermeasures' (2022) 17 Global Trade and Customs Journal.

¹⁵⁸ Military and Paramilitary Activities in and against Nicaragua (*Nicaragua v US*) (Merits) [1986] ICJ Rep. 4, para 186.

sanctions, rather than continuing to appeal to various extensions of the traditional jurisdictional principles. Having regard to the crucial necessity of strengthening the effectiveness and legitimacy of sanctions against Russia as a means of inducing it to comply with its international obligations vis-à-vis Ukraine, the time for such an explicit reliance has now come.

CONCLUSION

Against the background of the response of the international legal community to the Russian invasion of Ukraine in February 2022, this article has sought to examine the jurisdictional validity of extraterritorial sanctions (as imposed primarily by the US). The traditional bases of jurisdiction – territory, nationality, security – cannot justify the imposition of extraterritorial sanctions to third parties. Therefore, anti-evasion should be recognised as a new ground of jurisdiction to confer validity on a limited number of measures targeting transactions designed to circumvent lawful sanctions. This would provide the necessary legitimisation to many sanctions measures currently imposed against Russia in connection with its aggression against Ukraine. Given the dual purposes of governance and sovereignty underlying the doctrine of jurisdiction, the recognition of anti-evasion as a jurisdictional ground would not be an impermissible extension of the rules of jurisdiction.

The Forgotten Sector: Establishing the Need for Stronger Regulation of Livestock Emissions in the European Union – An Emergency Law Perspective

Katharina Neumann*

ABSTRACT

The current EU emission mitigation framework advances the narrative that CO₂ emissions are the predominant source of anthropogenic climate change. However, this approach fails to acknowledge the intense impact of non-CO₂ greenhouse gas emissions, primarily from animal agriculture, on global warming. This paper establishes the near exemption that livestock emissions are given within EU climate change policies, leading to inherently inadequate mitigation efforts. Although climate emergency declarations are continuously cited as efficient normative tools to overcome governmental complacency regarding climate policy, an investigation of the European Parliament's Climate Emergency Declaration suggests that the EU failed to exploit the potential of its climate emergency declaration as it retains a free pass for livestock emission within its current emissions mitigation framework. This paper proposes sectoral emergency intervention as an alternative to the current climate emergency declaration to stimulate necessary climate action in the livestock sector.

*BCL candidate at the University of Oxford specialising in Public International Law, LLB (Trinity College Dublin), The author would like to thank Orla Hughes and especially Professor Suryapratim Roy for their very helpful comments on an earlier draft of this article.

INTRODUCTION

Anthropogenic climate change, the concept that humans are causing the current shifts in atmospheric conditions by releasing large amounts of greenhouse gases (GHG),¹ is receiving growing attention as the earth's ecosystem approaches an irreversible tipping point. Many countries, organisations, and municipalities have declared climate emergencies in order to exploit the normative potential of emergency frameworks. The European Parliament declared a climate emergency in 2019, calling on 'the Commission, the Member States, and all global actors' to 'urgently take the concrete action needed to fight and contain this threat before it is too late'.²

Carbon dioxide (CO₂), methane (CH₄), and nitrous oxide (N₂O) are the three largest pollutants contributing to global warming, and reduction efforts are necessary to combat the climate emergency.³ Most non-CO₂ emissions (CH₄ and N₂O) stem from agricultural practices, specifically from livestock.⁴ A recent report from the environmental think-tank 'Worldwatch Institute' states that livestock emissions account for 51% of greenhouse gases. It found that a prior estimate from the Food and Agriculture Organization (FAO), which indicated that livestock accounted for 18% of GHG, overlooked some direct and indirect livestock emissions (including CO₂ emissions from livestock respiration, methane emissions, and emissions from clearing land to graze livestock and

¹ Stefan Rahmstorf, 'Anthropogenic Climate Change: Revisiting the Facts' in Ernesto Zedillo (eds), *Global Warming: Looking Beyond Kyoto* (Brookings Institution Press 2008) 35.

² Resolution on the Climate and Environment Emergency 2019/2930(RSP) <[https://oeil.secure.europarl.europa.eu/oeil/popups/ficheprocedure.do?lang=en&reference=2019/2930\(RSP\)](https://oeil.secure.europarl.europa.eu/oeil/popups/ficheprocedure.do?lang=en&reference=2019/2930(RSP))> accessed 21 February 2022.

³ Gunnar Myhre and others, 'Anthropogenic and Natural Radiative Forcing' in T. F. Stocker and others (eds), *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge University Press 2013) 659-740.

⁴ Hongmin Dong and others, 'Chapter-10: Emissions from Livestock and Manure Management' in Simon Eggleston and others (eds), *2006 IPCC Guidelines for National Greenhouse Gas Inventories, Volume 4 - Agriculture, Forestry and Other Land Use* (Institute for Global Environmental Strategies 2006) 10.68.

grow feed).⁵ Despite this, the livestock sector has been subject to less stringent climate change mitigation policies, with carbon emissions from fossil fuel extraction receiving far more attention.⁶ Although climate emergency declarations support an ‘all-hands-on-deck’ approach to combating climate change, even where such declarations are made, there continues to be a narrow focus on fossil fuel emissions. This paper argues for a sectoral emergency declaration in the livestock sector to end its unregulated position within emission mitigation schemes. The failure of the current EU climate emergency declaration to end livestock emissions’ free pass makes this approach a necessity.

This paper is divided into four parts. Section I seeks to conceptualise the European Climate Emergency Declaration by situating it in a broader emergency law framework. Section II establishes the exceptional position of animal agriculture with regard to its regulation within the EU’s emission mitigation framework as opposed to carbon emissions from other sectors. Section III explores general and institutional explanations for the failure of the European Parliament’s Emergency Declaration to encourage appropriate action. Finally, Section IV makes a case for a sectoral emergency declaration. It sets out three justifications for sectoral emergency intervention to end the failure to address livestock emissions, and highlights the potential of such an emergency declaration to reduce them.

SECTION I: EMERGENCY LAW CONCEPTUALISATIONS - THE EUROPEAN UNION’S CLIMATE EMERGENCY DECLARATION

The concept of ‘emergency’ and its operation invite different understandings of emergency regimes. This section clarifies key definitions and concepts as a basis for subsequent sections.

⁵ World Watch, ‘Livestock and Climate change’ (*World Watch*, 2019) <<https://awellfedworld.org/wp-content/uploads/Livestock-Climate-Change-Anhang-Goodland.pdf>> accessed 17 February 2023.

⁶ Bojana Bajželj and others, ‘Importance of Food-Demand Management for Climate Mitigation’ (2014) 4 *Nature Climate Change* 928.

1.1 Climate Change as an Emergency

Classifying phenomena as emergencies entails both a factual and value-based analysis, including both significant scientific evidence as well as proof of public crisis perceptions.

Compared with rapidly-forming emergencies such as the COVID-19 pandemic, climate change is best described as a ‘catastrophe in slow motion’.⁷ Nonetheless, its ramifications will be disastrous, threatening national and international security, as evidenced by the 2022 International Panel on Climate Change (IPCC) report. The report states that ‘beyond 2040 ... climate change will lead to numerous risks to natural and human systems’ including significant increases in ‘ill health and premature deaths’ and heightened risk of ‘displacement in all regions’. This leads to fragile states and regions where increased resource conflicts allow malign actors to thrive.⁸ For example, Drury and Olson have developed a model that statistically identifies a positive linkage between disaster severity and levels of political unrest.⁹ Nel and Righarts have identified that climate-related disasters increase the risk of conflict onset in the same year and the following year.¹⁰ Climate change in this sense is defined as a threat multiplier as the resulting resource scarcity multiplies inequalities, rebellion opportunities and incentives.

Although there is scientific evidence that if global warming increases by 1.5°C, there will be unavoidable increases in multiple climate hazards.¹¹ To stress the immense likelihood of irreversible future damage due to climate

⁷ Bruno Latour, ‘The Other State of Emergency’ (2016) 22 (3) *Social Identities* 228; Raymond T. Pierrehumbert, ‘Climate Change: A Catastrophe in Slow Motion’ (2006) 6 *Chicago Journal of International Law* 573.

⁸ Hans Pörtner and others, ‘Summary for Policymakers IPCC Sixth Assessment Report’ (IPCC, 2022) <https://report.ipcc.ch/ar6wg2/pdf/IPCC_AR6_WGII_SummaryForPolicymakers.pdf> accessed 22 March 2022.

⁹ Cooper A. Drury and Richard Stuart Olson, ‘Disasters and Political Unrest: An Empirical Investigation’ (1998) 6 (3) *Disasters and Political Unrest* 153.

¹⁰ Philip Nel and Marjolein Righarts, ‘Natural Disasters and the Risk of Violent Civil Conflict’ (2008) 52 (1) *International Studies Quarterly* 159.

¹¹ H. O. Pörtner and others (eds), ‘Summary for Policymakers’ in H. O. Pörtner and others (eds), *Climate Change 2022: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge University Press 2022).

change,¹² this paper accepts the precautionary principle as the central principle for framing climate change as an emergency. The precautionary principle is an integral part of international environmental law to close existing gaps and has been adopted by many domestic jurisdictions as well. It is set out in the Rio Declaration that ‘where there are threats of serious or irreversible damage, lack of full scientific certainty shall not be used as a reason for postponing cost-effective measures to prevent environmental degradation’.¹³ This is crucial for an issue such as climate change where it is impossible to identify all its potential consequences and how they will affect humanity. In light of the principle, ‘anticipation and/or prevention of harm notwithstanding an absence, deficit or ambivalence of knowledge suffices to find that an action is an emergency’.¹⁴ Thus, reference to precaution motivates action where repercussions are not yet visible or still unfolding, as it produces a causal connection between today’s science and tomorrow’s consequences. This is especially important due to the imminent danger and unforeseeable nature of the extreme threats that climate change might produce. Further, the 2021 UNDP Climate Vote Survey, including over half of the world population, found that 64% of voters believe climate change is a global emergency,¹⁵ proving that climate change is also a public crisis.

Thus, climate change possesses the constitutive features of an emergency,¹⁶ warranting the formulation of climate emergency declarations, which are contingent on the existence of a recognised state of emergency.¹⁷

¹² Susan Solomon and others, ‘Irreversible Climate Change due to Carbon Dioxide Emissions’ (2009) 106 Proceedings of the National Academy of Sciences 1704.

¹³ UN General Assembly, Rio Declaration on Environment and Development (*United Nations*, 1992) <https://www.un.org/en/development/desa/population/migration/generalassembly/docs/globalcompact/A_CONF.151_26_Vol.I_Declaration.pdf> accessed 17 February 2023.

¹⁴ Bruce Lindsay, ‘Climate of Exception: What might a Climate Emergency mean in Law?’ (2010) 38 (2) Federal Law Review 255, 271-273.

¹⁵ UNDP, ‘World’s largest survey of public opinion on climate change: a majority of people call for wide-ranging action’ (*United Nations Development Programme*, 2021) <<https://www.undp.org/press-releases/worlds-largest-survey-public-opinion-climate-change-majority-people-call-wide>> accessed 1 January 2022.

¹⁶ Jocelyn Stacey, ‘The Environmental Emergency and the Legality of Discretion in Environmental Law’ (2016) 52 Osgoode Hall LJ 985, 1027.

¹⁷ John Ferejohn and Pasquale Pasquino, ‘The Law of the Exception: A Typology of Emergency Powers’ (2004) 2 (2) International Journal of Constitutional Law 210.

1.2 Conceptualising Emergency Powers

Notions of emergency powers can be distinguished in their approach to the rule of law. Historically, under the neo-Roman model, emergencies have been linked to suspension of law.¹⁸ Introducing a dialogue between law and politics, Schmitt defines emergencies as states of exception where a sovereign power has absolute capacity to transcend the rule of law in the name of the public good.¹⁹ Agamben's concept of *exception-as-the-rule* goes even further,²⁰ indicating that emergencies induce the collapse of the defining relations between law and exception, suggesting that the exception autonomously defines the political.²¹ However, the suspension of the rule of law, even if only to a limited extent, permits abuse as actions remain unreviewable, potentially creating a super-constitutional emergency government (i.e., a governing body that exceeds the powers given to the executive under the constitution).²²

The EU Climate Emergency Declaration specifically states that 'no emergency should ever be used to erode democratic institutions or undermine fundamental rights ... and all measures will always be adopted through a democratic process'.²³ This indicates that the Declaration envisions an emergency regime that operates within the ordinary democratic process as opposed to being subject to extraordinary powers of the executive. Thus, when referring to emergency powers, this paper endorses a collaborative emergency regime co-produced by many,²⁴ defending the extension of the rule of law during emergencies,²⁵ and reinforcing the democratic legitimacy of emergency measures. This legislative approach to emergency law-making – conceptualising

¹⁸ David Dyzenhaus, 'Now the Machine Runs Itself: Carl Schmitt on Hobbes and Kelsen' (1994) 16 *Cardozo Law Review* 10.

¹⁹ Naomi Lazar, 'Must Exceptionalism Prove the Rule? An Angle on Emergency Government in the History of Political Thought' (2006) 34 (2) *Politics & Society* 245, 247.

²⁰ Giorgio Agamben, *State of Exception* (University of Chicago Press 2005) 9.

²¹ Jef Huysmans, 'The Jargon of Exception—On Schmitt, Agamben and the Absence of Political Society' (2008) 2 *International Political Sociology* 166, 172-176.

²² Sascha Müller, 'Turning Emergency Powers Inside Out: Are Extraordinary Powers Creeping into Ordinary Legislation' (2016) 18 *Flinders LJ* 295, 301.

²³ Resolution on the Climate and Environment Emergency (n 2).

²⁴ Roxanne Doty, 'States of Exception on the Mexico–US Border: Security' (2007) 1 (2) *International Political Sociology* 116.

²⁵ William Scheuerman, 'Rethinking Crisis Government' (2002) 9 *Constellations* 492; Lazar (n 19) 266.

emergency responses as acts of the legislature working within its normal competence – is imperative as it supports the EU political framework, which states that ‘the notion of emergency powers is not enshrined as a regulative ideal to cope with crises’.²⁶ Emergency legislation may delegate a great deal of special and temporary powers to the executive, but this is continuously monitored by the legislature²⁷ and subject to independent control mechanisms.²⁸ This allows for the establishment of a separate emergency legal system with rules, rights, and procedures, rather than making exceptions to the law. The model enacts new laws in an exceptional way through emergency discourse by incentivising the legislature to act promptly and forcefully.²⁹

However, the model bears the risk that by embedding extraordinary powers in ordinary legislation, they potentially become normalised. This may, for one, lead to common rule of law concerns as the emergency framework operates with reference to ‘exceptionalism’ and ‘urgency’, which may risk the equality of all citizens before the law or the prevention of arbitrary use of power. Further, normalisation may provoke a reduction of the normative force inherent in an emergency declaration.³⁰ This threat is especially dominant within a climate crisis that demands continuous, prolonged efforts, which cannot be managed by temporary emergency governance.³¹ This indicates that the conservative nature from which the model derives its legitimacy is lost. Further, legislatures are large and slow and may be unable or unwilling to quickly respond to crises.³²

1.3 The Normative Potential of Emergency Declarations

Climate Emergency Declarations are symbolic measures without legal force. However, by instituting emergency discourse, they mobilise social and economic resources and set a new status quo, indicating its normative potential to activate emergency powers.

²⁶ Christian Kreuder-Sonnen, ‘Does Europe Need an Emergency Constitution’ (2021) 71 (1) *Political Studies* 125.

²⁷ Ferejohn and Pasquino (n 17) 215.

²⁸ Müller (n 22) 303.

²⁹ Andrew Neal, ‘Normalization and Legislative Exceptionalism: Counterterrorist Lawmaking and Changing Times of Security Emergencies’ (2012) 6 *International Political Sociology* 273.

³⁰ Müller (n 22).

³¹ Ferejohn and Pasquino (n 17) 228.

³² *ibid* 219.

Declaring a climate emergency signals both high risk and high urgency,³³ conveying the need for exceptionally quick action outside the prevailing economic and socio-ecological norms.³⁴ The EU's Emergency Declaration adopted this narrative, highlighting 'immediate and ambitious action ... crucial to limiting global warming to 1.5°C,' involving 'citizens and all sectors of society and the economy'.³⁵ Thus, climate emergency declarations can contribute to significant shifts in public perceptions of climate change from a 'future risk' to a 'current crisis'.³⁶ Emergency framing may serve as a tool for rapid social mobilisation,³⁷ encouraging bespoke and innovative policymaking and overcoming government complacency.³⁸ There are estimates that by 2050, the result of these declarations, and consequential response plans, in the UK alone will reduce CO₂ emissions by approximately 2.5 billion tonnes of CO₂ equivalents.³⁹ Thus, emergency declarations are an important persuasive tool, directly influencing governance processes by demanding a radical, urgent mobilisation of economic and social resources at an intense level.⁴⁰

Further, climate emergency declarations entrench political positioning of climate action intentions.⁴¹ This may establish a new status quo, setting a 'psychologically powerful baseline against which future policy failures can be measured'.⁴² The support of influential authorities prompts people to

³³ Lindsay (n 14) 259.

³⁴ *ibid* 262; Patrick Hodder and Brian Martini, 'Climate Crisis? The Politics of Emergency Framing' (2009) 44 (36) *Economic and Political Weekly* 54.

³⁵ Resolution on the Climate and Environment Emergency (n 2).

³⁶ Lucy McHugh and others, 'Risk? Crisis? Emergency? Implications of the New Climate Emergency framing for Governance and Policy' (2021) 12 (6) *WIREs Climate Change* 3.

³⁷ Ben Anderson, 'Emergency Futures: Exception, Urgency, Interval, Hope' (2017) 65 (3) *The Sociological Review* 463.

³⁸ Arjen Boin and others, *The Politics of Crisis Management: Public Leadership under Pressure* (2nd edn, Cambridge University Press 2016) 65.

³⁹ Climate Emergency Declaration, 'Climate emergency declarations in 2,318 jurisdictions and local governments cover 1 billion citizens' (*Climate Emergency Declaration*, 2023) <<https://climateemergencydeclaration.org/climate-emergency-declarations-cover-15-million-citizens/>> accessed 21 February 2023.

⁴⁰ McHugh (n 36) 2.

⁴¹ Xira Ruiz-Campillo and others, 'Motivations and intended outcomes in local governments' declarations of climate emergency' (2021) 9(2) *Politics and Governance* 17, 22.

⁴² Arden Rowell and Josephine van Zeben, 'The New Status Quo of the Paris Agreement: The Psychological Impact of the 2-Degrees Aspiration' (2016) *European Journal of Risk Regulation* 49, 49.

consciously and subconsciously justify the new state of affairs.⁴³ Thus, emergency declarations may produce moral obligations to avoid generating additional harm in a situation deemed to require immediate mitigation, thereby inciting a material response to the emergency.⁴⁴

Despite this normative potential, it must be recalled that emergency declarations are not legally enforceable. As a result, the generated normative obligations must be translated into enforceable measures to take effective climate action.⁴⁵ By instigating emergency discourse, declarations may justify enacting new laws in an exceptional way to address the climate crisis. However, this depends on whether the generated normative force is powerful enough to overcome government complacency. Unfortunately, policymakers often react by formulating ‘placebo policies’,⁴⁶ instead of enacting essential ‘treatment policies,’ to address the root of the issue. This curtails the practical benefits of emergency declarations. ‘Placebo policies’ refer to any sort of action which has the sole aim of demonstrating that something is being done about the problem, even where the solution has no link to the actual problems at hand. An example of this is Australia’s ‘Future Fuels and Vehicles Strategy’ which addresses skyrocketing emissions in the transport sector with a technology-led approach, instead of committing to robust mitigation targets.⁴⁷ However, as a whole, these drawbacks can be addressed in a separate manner and do not deter from the normative power of emergency declarations in mobilising social and economic resources and setting a new status quo.

SECTION II: LIVESTOCK EMISSIONS’ EXCEPTIONAL STATUS - ANALYSING EU MITIGATION POLICIES

Livestock emissions are frequently overlooked in emission mitigation policies. Even the most celebrated international climate treaties, such as the

⁴³ *ibid.*

⁴⁴ *ibid* 51.

⁴⁵ Ruiz-Campillo (n 41) 20.

⁴⁶ Allen McConnell, ‘The use of placebo policies to escape from policy traps’ (2019) 27(7) *Journal of European Public Policy* 957, 958.

⁴⁷ Department of Industry, Science and Resources, ‘Future Fuels and Vehicles Strategy’ (*Department of Industry, Science and Resources* 2022) <<https://www.industry.gov.au/news/machinery-government-mog-changes-our-department-1-july-2022>> accessed 17 February 2023.

Paris Agreement, have adopted a singular approach to industrial emissions.⁴⁸ This section will reveal similar patterns within the EU emission mitigation policy. The EU made commitments to reduce agricultural emissions, proposing focused climate change instruments on agricultural activities, and to include agriculture within its Nationally Determined Contribution (NDC). However, the main instruments regulating agricultural activity, the Common Agricultural Policy (CAP) and the Effort Sharing Regulation (ESR), fail to address livestock emissions and are fundamentally unbalanced, prioritising productivity and farm revenue over environmental concerns.⁴⁹ It is acknowledged that there may be inherent obstacles to the regulation of livestock emissions such as their nature, which are from non-point sources and thus difficult to identify. Further, several European countries have large farming lobbies which have great influence over policymaking as they are important economic actors. Most importantly, there is a connection between the regulation of livestock emissions and food security concerns, as regulating agricultural practices may decrease production and thus limit the availability of food. However, as is argued in Section IV, this essay finds the current regulatory framework insufficient in spite of these concerns.

2.1 Common Agricultural Policy (CAP) — Failing to Incentivise Reduction

The CAP is the EU's main agricultural and farming subsidies programme. While it strives for 'sustainable management of natural resources and climate action'⁵⁰ its financing structure is a major obstacle to meeting necessary climate objectives as it supports a highly industrialised, intensive, and livestock-focused agricultural model. This sustainable management would be necessary to comply with the Treaty of the Functioning of the European Union (TFEU), which requires environmental protection measures in all EU policies.⁵¹

⁴⁸ Jonathan Verschuuren, 'Climate change and agriculture under the UNFCCC and related documents' in Mary Jane Angelo and Anél du Plessis (eds), *Research Handbook on Climate Change and Agricultural Law* (Edward Elgar Publishing 2017), 44-45.

⁴⁹ Guy Pe'er and others, 'A greener path for the EU Common Agricultural Policy' (2019) 365 (6452) *Science* 449, 450.

⁵⁰ European Commission, 'The CAP and Climate Change' (*European Commission*, 2019) <https://ec.europa.eu/info/food-farming-fisheries/sustainability/environmental-sustainability/climate-change_en> accessed 15 March 2022.

⁵¹ Case C-428/07 *Horvath* [2009] ECR I-06355 [29].

Agricultural sustainability is intended through cross-compliance mechanisms, green direct payment ('greening') schemes, and rural development policies (RDPs).⁵² These have been widely ineffective, however,⁵³ failing to include reduction targets in subsidy distribution and to limit livestock numbers or provide incentives to reduce them.⁵⁴ The cross-compliance scheme makes financial support conditional on maintaining certain environmental standards,⁵⁵ but makes no reference to livestock within these codes of conduct.⁵⁶ The greening scheme, which rewards farmers for preserving natural resources fails to acknowledge the potential role of livestock in carbon sequestration.⁵⁷ Moreover, the flexibility of the scheme induces Member States to implement greening schemes to minimise burdens and protect farmers.⁵⁸ This is an obstacle to meeting necessary climate objectives as, instead of preserving the status quo, farmers must step up their climate protection measures. Lastly, although specifically supporting the shift towards a low carbon and climate-resilient economy in agriculture, merely 2% of RDPs between 2014–2020 focused on innovative livestock management to reduce GHG.⁵⁹ These failures to recognise the harm of livestock emissions and incentivise their reduction indicate their exceptional position in EU emission mitigation policies.

⁵² European Commission, 'EU agriculture policy and climate change' (*European Commission*, 2020)

<[https://www.europarl.europa.eu/RegData/etudes/BRIE/2020/651922/EPRS_BRI\(2020\)651922_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2020/651922/EPRS_BRI(2020)651922_EN.pdf)> accessed 22 December 2021.

⁵³ Aris Konstantinidis, 'European Court of Auditors Special Report-Greening: a more complex income support scheme, not yet environmentally effective' (2017) <https://www.eca.europa.eu/Lists/ECADocuments/SR17_21/SR_GREENING_EN.pdf> 9 September 2021.

⁵⁴ *ibid* 12.

⁵⁵ Council Regulation (EC) 73/2009 establishing common rules for direct support schemes for farmers under the CAP [2009] OJ L30/16, Art 4.

⁵⁶ Regulation (EU) 1306/2013 Rules on cross-compliance pursuant to Article 93 [2013] OJ L347/549; Guy Pe'er and others, 'Is the CAP fit for purpose? An evidence-based fitness check assessment' (2017) German Centre for Integrative Biodiversity Research 4.

⁵⁷ Alliance Environment and Thünen Institute, 'Evaluation study of the payment for agricultural practices beneficial for the climate and the environment' (2017) <<https://op.europa.eu/en/publication-detail/-/publication/598b81ff-dfbc-11e7-9749-01aa75ed71a1>> accessed 6 October 2021.

⁵⁸ Konstantinidis (n 53) 36.

⁵⁹ European Commission, Overview of the 118 RDPs 2014-2020 (2018) <https://ec.europa.eu/agriculture/sites/agriculture/files/rural-development-2014-2020/country-files/common/rdplist_en.pdf> accessed 6 October 2021.

The inconsistencies of this legislative framework become clear when compared to emission policies concerning other sectors, such as energy. Although these policies require improvement as subsidies for fossil fuel-powered energy remain prevalent,⁶⁰ Member States have acknowledged the harmfulness of these energy sources; agreed to phase out environmentally harmful subsidies⁶¹ and recognised that European Investment Bank (EIB) investments must ‘no longer support or de-risk unsustainable investments such as fossil fuels’.⁶² No such commitments have been made regarding livestock emission reductions. The comprehensive 10-year *Farm to Fork Strategy*, despite recognising the unsustainability of current livestock production and consumption, lacks commitment to livestock emissions.⁶³ The lack of commitment is illustrated by unambitious targets and loopholes which are exploited by self-interested states to protect the status quo for farmers rather than advancing climate change objectives. Ultimately, the EU adopts a singular focus on incentivising the reduction of emissions from other sectors, granting immunity to livestock emissions.

⁶⁰ Julian Nowag and others, ‘Phasing out fossil fuel subsidies in the EU? Exploring the role of state aid rules’ (2021) 21(8) *Climate Policy* 1037.

⁶¹ European Commission, ‘Second Report on the State of the Energy Union’ (*European Commission*, 2017) <https://ec.europa.eu/commission/sites/beta-political/files/2nd-report-state-energy-union_en.pdf> accessed 7 October 2021; Council Decision 2010/787/EU on State aid to facilitate the closure of uncompetitive coal mines [2010] OJ L336; European Commission, ‘A roadmap for moving to a competitive low carbon economy in 2050’, 10 (*European Commission*, 2011) <<http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52011DC0571&from=EN>> accessed 16 October 2021.

⁶² European Commission, ‘Financing a Sustainable European Economy’ (*European Commission*, 2017)

<https://ec.europa.eu/info/sites/info/files/170713-sustainable-finance-report_en.pdf> accessed 8 October 2021.

⁶³ European Commission, *Farm to Fork Strategy* (*European Commission*, 2020), 7–8 <<https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020DC0381>> accessed 8 October 2021.

2.2 Effort Sharing Regulation (ESR) – Failing to Regulate Appropriate Reduction

The ESR is a governance tool, providing an overarching reduction target for the EU and assigning individual Member State targets.⁶⁴ The Regulation replaces the Effort Sharing Decision (ESD),⁶⁵ which regulated but did not limit agricultural emissions.⁶⁶ It supplements the Land Use and Forestry Regulation (LULUCF) which sets ‘no-debit’ targets for net emissions/removals from agricultural land use.⁶⁷ The ESR implements binding, enforceable reduction targets for non-CO₂ emissions, implemented through national law, which allows for some flexibility.⁶⁸ Addressing emissions from sources not covered by the EU emissions trading system (ETS), the ESR aims to reduce emissions by 40% by 2030 from 1990 levels.

Although the ESR advances mitigation efforts in the agricultural sector, reluctance to target some emissions is noticeable. As reduction targets are aggregated to cover all non-ETS sectors, Member States can focus on reducing emissions in some sectors while avoiding others. Further, the Regulation includes two flexibility mechanisms to relieve the ‘pressure to reduce agricultural emissions’.⁶⁹ The LULUCF flexibility allows Member States to use LULUCF credits to offset emissions in the ESR sector, and the ETS flexibility entitles Member States to use ETS allowances up to the equivalent of 2% or 4%

⁶⁴ Regulation (EU) 2018/842 on binding annual greenhouse gas emission reductions by Member States from 2021 to 2030 contributing to climate action to meet commitments under the Paris Agreement [2018] OJ L156.

⁶⁵ Decision (EU) No 406/2009/EC of the European Parliament and of the Council of 23 April 2009 on the effort of Member States to reduce their greenhouse gas emissions to meet the Community’s greenhouse gas emission reduction commitments up to 2020 [2009] OJ L140.

⁶⁶ *ibid.*

⁶⁷ Regulation (EU) 2018/841 on the inclusion of greenhouse gas emissions and removals from land use, land use change and forestry in the 2030 climate and energy framework [2018] OJ L156.

⁶⁸ Marjan Peeters, ‘The continued effort sharing approach in EU climate law: Binding targets, challenging enforcement?’ (2020) 29(2) *Review of European, Comparative & International Environmental Law* 201, 201.

⁶⁹ Regulation (EU) 2018/842 (n 64).

of their 2005 non-ETS emissions.⁷⁰ This practically leads to the agricultural sector merely having to reduce emissions by 6% if the flexibility is fully exercised. Thus, excessive flexibility disincentivises mitigation action in the agricultural sector,⁷¹ prompting agriculture to only contribute 1% of the 11% ETS reduction effort between 2005-2018.⁷² The Regulation also fails to reference livestock emissions specifically, despite them being the main share of agricultural emissions.⁷³ Ultimately, the Regulation's indifferent approach towards regulating livestock emission reduction further indicates their free pass within EU emission mitigation frameworks.

Further, the ESR is an arguably weaker instrument than the market-based ETS, addressing emissions from other sectors. The ETS sets a determinative price on carbon, delivering a predictable reduction pathway through its declining cap, sending long-term signals for businesses and investments. Contrarily, the ESR faces issues in targeting specific emissions and enforcing such targets. Thus, the inability of the ESR to mandate emission reduction as effectively as the ETS further confirms the focus of emission mitigation policies on mandating emission reduction from other sectors.

Ultimately, the scale of GHG reductions required to combat the climate emergency exempts no sector,⁷⁴ as 'full mitigation potential ... will only be realised if agricultural emissions are included in mainstream climate policy'.⁷⁵ In July 2021, the European Commission published its environmental package 'Fit for 55,' which presented a suitable opportunity to end livestock's free pass

⁷⁰ IFOAM Organic Europe, 'IFOAM EU position on the Effort Sharing Regulation and Land Use, Land Use Change and Forestry' (2017) 2
<https://www.organicseurope.bio/content/uploads/2020/06/ifoameu_policy_euclimatenergy2030_positionpaper_201702.pdf?dd> accessed 4 March 2022.

⁷¹ *ibid* 5.

⁷² European Environment Agency, 'National action across all sectors needed to reach greenhouse gas Effort Sharing targets' (*EEA*, 2020)
<<https://www.eea.europa.eu/themes/climate/trends-and-projections-in-europe/national-action-across-all-sectors>> accessed 4 March 2022.

⁷³ IFOAM (n 70) 7.

⁷⁴ Meryl Richards and others, 'National contributions to climate change mitigation from agriculture: Allocating a global target' (2018) 18(10) *Climate Policy* 1283.

⁷⁵ Margot Hurlbert and others, 'Chapter-7: Risk management and decision making in relation to sustainable development' (2019) 676
<https://www.ipcc.ch/site/assets/uploads/sites/4/2019/11/10_Chapter-7.pdf> accessed 4 March 2022.

and exploit the force of the Parliament's Declaration. However, the package does not entail the needed reforms. The CAP continues a business-as-usual approach to farming, merely providing for more flexibility through national CAP strategic plans and slightly more ambitious targets and funding to the existing CAP mechanisms.⁷⁶ Regarding the ESR, neither the loopholes concerning the flexibilities, nor the aggregated emission target approach, have been revised.⁷⁷ Overall, agricultural emissions have barely declined since 2005.⁷⁸ Future predictions taking into account implemented policies also show that no reduction is expected. Thus, as mitigation of CO₂ emissions from other industries is intensifying, this consistent level of agricultural emissions indicates the continuing privilege afforded to livestock emissions, despite their enormous climate impact. Any ambitious mitigation policy to fight climate change must make it a priority to change this. Importantly, this paper does not suggest that livestock emissions should be regulated instead of emissions from industry, but advances the argument that the current climate emergency requires an 'all-hands-on-deck' approach which must include livestock emission reductions.

SECTION III: THE FAILURE OF THE EUROPEAN PARLIAMENT'S CLIMATE EMERGENCY DECLARATION - POTENTIAL EXPLANATIONS

Only 11% of climate emergency declarations materialise into institutional change, and the European Parliament's Climate Emergency Declaration is unlikely to change this bleak statistic.⁷⁹ This section explores the EU institutional structure and general emergency framing issues as possible explanations for the failure to exploit the value of the European Parliament's Declaration to alter the exceptional position of livestock within its emissions mitigation policy.

⁷⁶ Margarethe Scheffler and Kirsten Wiegmann, 'Improving the contribution of the Common Agricultural Policy to climate change mitigation' (2021) German Watch.

⁷⁷ European Commission Proposal for a Regulation of the European Parliament and the Council amending Regulation (EU) 2018/842 on binding annual greenhouse gas emission reductions by Member States from 2021 to 2030 contributing to climate action to meet commitments under the Paris Agreement [2018] COM/2021/555.

⁷⁸ EEA (n 72).

⁷⁹ Ruiz-Campillo (n 41) 25.

3.1 The EU Institutional Structure as an Explanation for the Declaration's Lack of Success

i. Fragmented Legislative Power

A prominent feature of the legislative emergency model is that the legislature plays a fundamental role both in recognising an emergency and creating powers to deal with it.⁸⁰ In the EU institutional framework, legislative power is shared between the Parliament, the Council, and the Commission. While the right to legislative initiative lies with the Commission, the Council and the Parliament reject or amend Commission proposals under the ordinary legislative procedure, which has applied to environmental policymaking since the Amsterdam Treaty,⁸¹ and the CAP since the Lisbon Treaty.⁸² Given that the EU legislative process is more diffused than in national systems, it is difficult to formulate policy as urgently and forcefully as required for emergency intervention.⁸³ Thus, the EU institutional framework is ill-suited to employ the legislative emergency model as its disconnected legislative structure makes it almost impossible to reach a compromise on creating the powers necessary to deal with an emergency. Such powers mostly include the ability to produce emergency legislation.

This is also seen in the EU's approach of dealing with the financial crisis. As the EU had no adequate institutional mechanisms in place to deal with the situation, its response took the form of emergency politics characteristic of politics beyond the nation-state, resulting in an unregulated framework which exceeded the democratic political order (exemplified by expanding European Central Bank (ECB) competencies outside the treaties).⁸⁴ There was a need for fiscal integration to support the common currency but implementation pathways were blocked by 'the constraining dissensus among European publics',⁸⁵ creating several functional gaps in the EU authority structure. To fill

⁸⁰ Ferejohn (n 17) 217.

⁸¹ Paul Craig and Gráinne de Búrca, *EU Law: Text, Cases, and Materials* (7th edn, Oxford University Press 2020) 113-17.

⁸² Francisco Torres, 'How efficient is joint decision-making in the EU? Environmental policies and the co-decision procedure' (2003) 38(6) *Intereconomics* 312, 321.

⁸³ *ibid* 319.

⁸⁴ Kreuder-Sonne (n 26) 128-129.

⁸⁵ *ibid* 126.

these gaps and circumvent public dissent, decision-makers resorted to emergency measures outside the EU legal framework. They entrusted political interventions to the politically-independent and democratically-unchecked ECB.⁸⁶ Thus, during the financial crisis, the EU did not opt for a legislative emergency framework, as is suggested in this paper, but rather for democratically illegitimate and unregulated emergency powers by the ECB. With regard to climate action in the livestock sector, the lack of institutional mechanisms has so far led to legislative inaction, as the diffused system requires the political agreement of a variety of actors.

ii. Conflicting Interest of Legislative Organs

The condition for a functioning legislative emergency model is that the legislature must be willing to enact statutes conceding new, temporary powers to the executive: ‘The normal circumstances of jealousy or rivalry between governmental departments (must) be overridden during emergencies’.⁸⁷ However, as the composition and function of different EU organs can often lead to clashes in interest there is a strong rivalry between them. While there are instances where EU-wide interests will cohere with individual Member State interests, requisite majorities may be difficult to generate on topics such as livestock, which affect some states such as Spain and Italy more than others. Thus, in general, the *supranational*⁸⁸ Commission and the European Parliament (EP), in which representatives act in the interests of the Union, prefer furthering EU interests, whereas in the *intergovernmental*⁸⁹ Council, Member States’ national interests dominate as decisions are made by the relevant ministers in their capacity as state representatives.⁹⁰

Intergovernmental and supranational interests clash not only in legislative-executive power distribution but also during the ordinary legislative procedure, which is problematic as the procedure does not ‘create truly co-equal

⁸⁶ *ibid.*

⁸⁷ Ferejohn (n 17) 218.

⁸⁸ EU policy outcomes are results of bargaining by institutions which are not merely agents for national governments.

⁸⁹ EU policy outcomes are results of interstate bargaining among states pursuing their own national interests.

⁹⁰ Raj Chari and Sylvia Kritzing, *Understanding EU policy making* (Pluto 2006) 80.

branches in terms of relative policy influence'.⁹¹ The co-decision procedure has had positive impacts on reaching a compromise in favour of the EP within the specific context of environmental policy-making, and the EP has been increasingly able to shift policy outcomes since the introduction of the procedure towards its preferences. However, it still cannot reliably ensure outcomes that are closer to its own preferences than those of the Council, remaining heavily contingent on preference congruence. This establishes the Council as the dominant EU policy actor.⁹² Thus, intergovernmental interests persist in environmental policymaking, hindering ambitious climate and agricultural policies. However, a caveat must be introduced. Although the EP has been credited as a climate champion, mostly due to its influential and forward-looking environmental committee and its institutional location in the legislative process, its activism is highly dependent on its composition. The EP will be more inclined to pursue a progressive climate agenda if the majority of seats are occupied by parties that support active climate action.⁹³ While an empirical study has found that the EP's amendments regarding environmental protection are often unambitious and that it may have lost its teeth in environmental championing, the EP still seeks to advance the environmental agenda more than other EU organs.⁹⁴ The supranational Parliament, which is more inclined to adopt stronger climate policies regardless of differentiated national resistances, is stifled by the Council protecting national interests.⁹⁵ Thus, the lack of will to reduce emissions in the livestock sector may have arisen due to an impasse within the EU institutional structure.

The rejection of industry-related policies in order to protect national interests is exemplified in negotiations on the most recent CAP reform. The Policy ultimately adopted a voluntary approach to capping payments above €100,000/year, as preferred by the Council, although the Parliament was

⁹¹ Craig and de Búrca (n 81) 113-17.

⁹² Amie Kreppel, 'Bicameralism and the balance of power in EU legislative politics' (2018) 24(1) *The Journal of Legislative Studies* 11, 28.

⁹³ Charlotte Burns and Neil Carter, 'Is Co-decision Good for the Environment? An Analysis of the European Parliament's Green Credentials' (2010) 58 *Political Studies* 123, 137.

⁹⁴ Kreuder-Sonne (n 26) 142.

⁹⁵ Uwe Puetter, 'Europe's deliberative intergovernmentalism: the role of the Council and European Council in EU economic governance' (2012) 19(2) *Journal of European Public Policy* 161, Torres (n 82) 320.

defending mandatory application.⁹⁶ The dilution of crop rotation policies, by considering crop diversification as an equivalent in line with Council's policy preferences, is another example. Opposed to this, the Parliament pushed for additional safeguards prohibiting monocultures.⁹⁷ Crop diversification is important as it provides for more secure, stable wildlife habitats and increased resilience to climate change. Replacing crop rotation policies with crop diversification as an equivalent is detrimental as crop diversification cannot provide the same benefits as crop rotation. This highlights how the Parliament's ambitious climate policies are regularly undermined within the EU legislative process to protect Member States' interests. The trend is further apparent in other industries, exemplified by the Council continuously blocking a Commission proposal adopted by the Parliament to achieve climate neutrality by 2050 due to concerns that this would hurt nuclear and coal-dependent economies.⁹⁸

The dominance of Member State interests in EU policymaking explains why the EP's Declaration has not yet incited appropriate action regarding livestock emission mitigation. Considering the tension between intergovernmental and supranational organs of the EU, it is unlikely that the legislative emergency model's condition of non-rivalry between legislative and executive organs is currently satisfied within the EU. The fact remains that 'the EU is still far from resembling the hierarchically integrated state structures where constitutional provisions to accommodate emergency powers are commonplace'.⁹⁹

⁹⁶ Matteo Metta and Oliver Moore, 'Last Week of CAP Negotiations: What's the Deal?' (*Heinrich-Böll-Stiftung*, 2021) <<https://eu.boell.org/en/2021/05/25/last-week-cap-negotiations-whats-deal>> accessed 3 March 2022.

⁹⁷ *ibid.*

⁹⁸ European Council, 'European climate law: Council and Parliament reach provisional agreement' (*European Council*, 2021) <<https://www.consilium.europa.eu/en/press/press-releases/2021/05/05/european-climate-law-council-and-parliament-reach-provisional-agreement/>> accessed 18 February 2022.

⁹⁹ Kreuder-Sonne (n 26) 128.

3.2 Emergency Framing Issues as an Explanation for the Declaration's Lack of Success

Aside from the EU's institutional structure, emergency framing itself might explain why the Parliament's Declaration has not incited adequate climate action in the livestock sector. This is due to several reasons:

- (i) The narrative of 'catastrophe' communicating worry and fear, as well as implying a deadline in relation to a closing window,¹⁰⁰ might generate polarised actions because of expressed hopelessness,¹⁰¹ or even discourage efforts.
- (ii) The climate crisis is a complex collective action problem, requiring coordinated widespread systemic change across an intergenerational timespan. Thus, it does not fit the fundamentally 'conservative' nature of emergency frameworks. Emergency powers are inherently conservative as they are temporary measures aimed at the restoration of previous conditions. However, it is difficult to conceptualise the point at which the climate is restored to its 'original' position, hindering the application of emergency frameworks generally.¹⁰²
- (iii) The meanings of 'climate emergency declarations' are ambivalent, and they are described as an 'empty signifier' in that the multiple meanings ascribed to them reinforce different ideological projects.¹⁰³ The generated normative force may be inadequate to overcome the issues of the crisis, but rather lead to previously discussed placebo policies, such as progressive mitigation targets without implementation mechanisms and a general gap between declaration and implementation.

¹⁰⁰ Shinichiro Asayama and others, 'Why setting a climate deadline is dangerous' (2019) 9 *Nature Climate Change* 570, 571.

¹⁰¹ Andrew Wilson and Ben Orlove 'What do we mean when we say climate change is urgent?' (2019) *Colombia Academic Commons* <<https://academiccommons.columbia.edu/doi/10.7916/d8-b7cd-4136>> accessed 24 March 2022.

¹⁰² Ferejohn and Pasquino (n 17) 219.

¹⁰³ Chris Methmann, 'Climate protection' as empty signifier: A discourse theoretical perspective on climate mainstreaming in world politics' (2010) 39(2) *Journal of International Studies* 345, 348-9.

Ultimately, as it is proposed that ending livestock emissions' free pass must be prioritised in an appropriate mitigation policy, retaining the special position of these emissions in the EU indicates the failure of the EU Climate Emergency Declaration to incite action.¹⁰⁴ Emergency declarations are merely the first of several steps to initiate climate emergency mobilisation. However, the EU Declaration has been unable to break government complacency due to EU-specific institutional issues and more conceptual problems regarding emergency declarations.

SECTION IV: THE WAYS FORWARD - THE CASE FOR A SECTORAL EMERGENCY DECLARATION

Having established that the current framework regarding livestock emissions is inadequate despite the Parliament's Climate Emergency Declaration, this section proposes the declaration of sectoral emergency in the livestock sector as an alternative.

4.1 Justifying a Sectoral Emergency Law Response to Fight Livestock Emissions' Exceptional Status

Before dealing with the potential of a sectoral EU emergency declaration for the livestock industry, such a targeted intervention must be justified. It is argued that livestock emissions' significant reduction potential, associated food security concerns, and the principle of equal treatment should trigger the intervention of emergency law within the livestock industry.

¹⁰⁴ Mark Chou, 'Australian local governments and climate emergency declarations: Reviewing local government practice' (2020) *Australian Journal of Public Administration* 607, 618; Ruiz-Campillo (n 41) 17.

i. Reduction Efficiency

Many EU instruments aim to cut emissions where it is cheapest and most practicable,¹⁰⁵ advancing an efficiency narrative for reduction. The objective of this efficiency analysis in law is the value-maximisation of normative resources, requiring mitigation policies that maximise reduction capability.¹⁰⁶ GHG from livestock have considerable mitigation potential, both because it is currently not being actively reduced, and because of the nature of these emissions. Their inclusion would arguably reflect a more *efficient* reduction policy.

First, the impact of livestock emissions is currently not accounted for, thus suggesting capacity for massive reductions. Agriculture is a significant contributor to global warming, generating around half of all anthropogenic CH₄ emissions and around three-quarters of N₂O,¹⁰⁷ but has so far been disregarded in mitigation schemes. Livestock is the leading source of non-CO₂ GHGs,¹⁰⁸ indicating that animal agriculture's temperature impact accounts for at least 23%, without including emissions related to feed production, land-use change, fertiliser use, energy, and transport.¹⁰⁹ Although non-CO₂ GHG emissions from agriculture have declined by 24% between 1990 and 2012, this relative change is

¹⁰⁵ Examples include the ETS and emissions caps dependent on member states. This is also the rationale behind the European Emission trading scheme. See: European Commission, 'Directive 2003/87/EC establishing a system for greenhouse gas emission allowance trading within the Union' (*European Commission*, 2003) <<https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32003L0087>> accessed 18 February 2023; European Council and Parliament Regulation (EU) 2018/842 of the European Parliament and of the Council of 30 May 2018 on binding annual greenhouse gas emission reductions by Member States from 2021 to 2030 contributing to climate action to meet commitments under the Paris Agreement and amending Regulation (EU) No 525/2013. For example, Article 4 provides that 'Each Member State shall, in 2030, limit its greenhouse gas emissions at least by the percentage set for that Member State in Annex I in relation to its greenhouse gas emissions in 2005, determined pursuant to paragraph 3 of this Article.'

¹⁰⁶ Robert Cooter, 'Liberty, Efficiency, and Law' (1987) 50(4) *Law and Contemporary Problems* 141, 142.

¹⁰⁷ Cheikh Mbow and others, 'Chapter-5: Food security' (IPCC, 2019) <https://www.ipcc.ch/site/assets/uploads/sites/4/2020/02/IPCCJ7230-Land_SM5_200226.pdf> accessed 4 March 2022.

¹⁰⁸ Helen Harwatt, 'Including animal to plant protein shifts in climate change mitigation policy: a proposed three-step strategy' (2018) 19(5) *Climate Policy* 533, 535.

¹⁰⁹ Andy Reisinger and Harry Clark, 'How much do direct livestock emissions actually contribute to global warming?' (2017) 24 *Global Change Biology* 1749, 1750-51.

lower than in sectors such as waste and industrial processes,¹¹⁰ suggesting a possibility for further reduction.

Second, the distinct nature of non-CO₂ GHG demonstrates reduction potential. CH₄ has a short atmospheric lifetime but potent warming impact as opposed to CO₂ which slowly warms the planet over time. Thus, reductions may be achieved more rapidly, and effects will be observed more quickly compared to CO₂.¹¹¹ A large share of non-CO₂ mitigation can be realised at relatively low costs through land-use change, land management, and livestock management.¹¹² These steps are essential to meet higher emission targets and current net-zero targets.¹¹³ Thus, livestock emissions' inclusion in mitigation schemes allows for efficient reduction, complementing current CO₂ mitigation.

Ultimately, the capacity and nature of non-CO₂ GHG indicate that policies targeting non-CO₂ emission from livestock optimise value-maximisation of resources. Thus, cost-efficiency analysis provides a justification for targeted intervention of emergency law within the livestock sector to mitigate emission levels.

ii. Food Security Concerns

The right to food is a protected human right under Article 11(1) of the International Covenant of Economic, Social and Cultural Rights (ICESCR). However, climate change has slowed agricultural productivity over the past 50 years.¹¹⁴ The 2022 IPCC report clarifies that climate change will increase the

¹¹⁰ Ignacio Domínguez and Thomas Fellmann, 'The Need for Comprehensive Climate Change Mitigation Policies in European Agriculture' (2015) 14(1) *EuroChoices* 11, 11; Scheffler and Wiegmann (n 76) 2.

¹¹¹ Harwatt (n 108) 3; John Lynch and others, 'Agriculture's Contribution to Climate Change and Role in Mitigation Is Distinct From Predominantly Fossil CO₂-Emitting Sectors' (2021) *Frontiers in Sustainable Food Systems* 2; Peter Smith and others, 'Chapter-11: Agriculture, Forestry and Other Land Use (AFOLU)' in Peter Smith and others, (eds), *Working Group III contribution to the IPCC 5th Assessment Report "Climate Change 2014: Mitigation of Climate Change"* (IPCC 2014).

¹¹² *ibid.*

¹¹³ Mathijs Harmsen and others, 'Non-CO₂ greenhouse gases: the underrepresented, complex side of the climate challenge' (2020) 17(3) *Journal of Integrative Environmental Sciences* 3.

¹¹⁴ Pörtner (n 8) 10.

number of floods, droughts and heatwaves, putting pressure on our food systems and undermining food security.¹¹⁵ Food security concerns are further aggravated by population growth and changing preferences towards more emission-heavy diets, which worsen climate change. Currently, global meat, egg, and milk production are responsible for nearly one-fifth of human-induced GHG, and trends indicate that the livestock sector might contribute up to 70% of the GHG emission limit in 2050.¹¹⁶ Thus, the human right to food security is increasingly endangered by not only climate change itself but unsustainable agricultural practices and emission-heavy diets.

Food insecurity provides greater support for intervention in the livestock sector to address the issue.¹¹⁷ Apart from employing more sustainable agricultural techniques, incentivising shifts from animal to plant-sourced protein is effective in addressing both climate and food security concerns. Plant-based diets require fewer natural resources including land, nitrogen, phosphorus, water, and energy, resulting in lower GHG emissions and more space for carbon sinks and greater food availability.¹¹⁸ Thus, the growing concern surrounding the human right to food justifies the pointed emergency intervention in the livestock sector, as structural changes in the sector may not only decrease the food systems' climate impacts, but also alleviate fears of food insecurity.

iii. Equal Treatment

Under EU law, the general principle of equal treatment requires that comparable situations are not treated differently unless differentiation is objectively justified,¹¹⁹ adopting the Aristotelian approach that those in similar situations should be treated alike.¹²⁰ As non-CO₂ GHG from livestock have the

¹¹⁵ *ibid* 14-15.

¹¹⁶ Pierre Gerber and others, 'Tackling climate change through livestock—A global assessment of emissions and mitigation opportunities' (2013) UNFAO <<http://www.fao.org/3/8d293990-ea82-5cc7-83c6-8c6f461627de/i3437e.pdf>> accessed 7 February 2022.

¹¹⁷ Pörtner (n 8) 24.

¹¹⁸ Harwatt (n 108) 6.

¹¹⁹ T-704/14 *Marine Harvest ASA v European Commission* [2017] ECLI:EU:T:2017:753 [207].

¹²⁰ Dagmar Schiek, 'Torn between Arithmetic and Substantive Equality? Perspectives on Equality in German Labour Law' (2002) 18 *International Journal of Comparative Labour Law and Industrial Relations* 149, 150; T-10/93 *A v Commission* [1994] ECR II-179 [42].

same destructive effects on the climate system as CO₂, equal treatment requires their effective inclusion in mitigation schemes.

The *Arcelor* case exemplifies the application of this principle to a specific sector and its analysis can be applied to the livestock sector.¹²¹ The case concerned an action brought by Arcelor, a steel producer, challenging the validity of the Emissions Trading Directive. Arcelor claimed that certain articles of the Directive violate the principle of equal treatment. The company pointed out that the ETS includes the steel sector but not competing industries, including non-ferrous metal and chemical product sectors, despite their responsibility for comparable CO₂ emissions.¹²² The General Court explained that all GHG emissions contribute to climate change, and thus all sectors of the economy which emit them are comparable. The quantity of CO₂ emitted by each sector is immaterial for assessing comparability, indicating that the steel, chemical and non-ferrous metal sectors are in a comparable position.¹²³ However, the Court dismissed the equal treatment claim, holding that disparate treatment could be justified as the inclusion of non-ferrous metal and chemical product sectors would increase administrative burdens and disrupt ETS functioning.¹²⁴

Applying these principles to the livestock sector indicates that the sector's current free pass might breach equal treatment. *Arcelor* clarified that all emissions of greenhouse gases are liable to contribute to dangerous interference with the climate system and are thus comparable. Furthermore, the quantity of GHG emitted by a sector is not essential for assessing comparability, indicating that livestock emissions should not be distinguished from emissions from other sectors unless there are objective justifications. The *Arcelor* administrative burden justification has already been rebutted in the previous section, as there is value maximisation in allocating more administrative resources towards reducing emissions in the livestock sector. Food security concerns are also dismissed as a justification, and it is submitted that there are no further justifications for

¹²¹ C-127/07 *Société Arcelor Atlantique et Lorraine v Premier ministre* [2008] ECR I-09895.

¹²² *ibid* [162].

¹²³ *ibid* [34]-[38]

¹²⁴ *ibid* [59]-[65] cases concerning agriculture: C-96/03 and C-97/03 *Tempelman and van Schaijk* [2005] ECR I-1895 [48] and C-504/04 *Agrarproduktion Staebelow* [2006] ECR I-679 [37].

disparate treatment.¹²⁵ Thus, the intervention of emergency measures in this sector is justified to rectify the current situation, which illegally differentiates emissions reductions efforts between sectors, against EU law.¹²⁶

4.2 The Potential of a Sectoral Emergency Declaration to end Livestock Emissions' Free Pass

The potential of a sectoral emergency declaration to end the free pass of livestock emissions is derived from increased normative force and attribution of responsibility. A sectoral declaration limits the reach of the necessary 'restoration', thereby making an emergency framework more feasible. However, it must be submitted that the constraints posed by the EU institutional framework will continue to inhibit appropriate livestock emissions mitigation.

i. Increasing Normative Force

A sectoral emergency declaration will generate greater normative force to trigger emergency powers, incentivising climate action.¹²⁷ Pointed intervention shifts attention to livestock emissions, indicating that a sectoral declaration is not an 'empty signifier'. Its specific focus on the livestock sector communicates a communal ideological project, instigating a normative force for climate action potentially strong enough to overcome government complacency. This force is further strengthened by minimising the narrative of 'catastrophe', addressing a single sector as opposed to creating panic within the whole economy. Through this, such a declaration makes a *clearer* statement through the 'expressive function of the law'.¹²⁸ As Sunstein argues, such an expressive function means that an appropriately-framed law influences and pushes social norms to reconstruct existing norms and change the social meaning of action

¹²⁵ cf C-352/98 *Bergaderm and Goupil v Commission* [2000] ECR I-5291 [39] and C-198/03 *Commission v CEVA and Pfizer* [2005] ECR I-6357 [61].

¹²⁶ cf C-17/61 and C-20/61 *Klückner-Werke and Hoesch v High Authority* [1962] ECR 325 and C-250/83 *Finsider v Commission* [1985] ECR 131 [8]; C-462/99 *Connect Austria* [2003] ECR I-5197 [115].

¹²⁷ Daniel Bodansky, 'International Sectoral Agreements in a Post-2012 Climate Framework' (2007) Pew Center on Global Climate Change Working Paper <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1028187> accessed 4 February 2022.

¹²⁸ Cass Sunstein, 'On the expressive function of the law' (1999) 144 *University of Pennsylvania Law Review* 2021, 2024.

through a legal statement about appropriate behaviour.¹²⁹ As a sectoral declaration increases normative force, it arguably becomes *more* appropriate, and will thus have a greater effect on existing norms regarding livestock emission reduction and change the social meaning of livestock pollution more drastically.¹³⁰ For example, the entirely normative framework of the Paris agreement, including the 2°C temperature goal, the time target of peaking emissions in the latter half of the century, and the imposition of mitigation obligations on every state, has led to enforceable domestic laws adopting these norms.¹³¹

By making livestock pollution socially undesirable, such a declaration assists in bypassing EU institutional constraints. It generates a norm to which the society is committed and by which it should be legally bound,¹³² inducing the construction of more appropriate emission reduction policies in the livestock sector. Setting a uniform baseline by reference to such norms improves cooperation between institutional actors, holding them to the same normative standard and thus allowing the creation of emergency legislation through the ordinary legislative process. Overall, ‘transnational politics of emergency’ would not be ‘largely unregulated’ but rooted in social norms created by the declaration,¹³³ illustrating that it might be a viable alternative to activate emergency powers to deal with emission reduction. Importantly, this paper does not argue that emission reduction should be restricted to the livestock sector, but that the sector in particular would benefit from a sector-specific emergency declaration.

ii. Attributing Responsibility

Rather than a general emergency declaration involving ‘all sectors of society and the economy’,¹³⁴ a sectoral emergency declaration more expressly identifies responsible actors. While recognising individuals and EU decision-makers as part of the problem, it specifically shifts attention to those facilitating livestock emissions. Thus, such a declaration addresses the central

¹²⁹ *ibid* 2031.

¹³⁰ *ibid* 2032.

¹³¹ Examples include the German and Dutch Climate Change Act.

¹³² Sunstein (n 128) 2027.

¹³³ Kreuder-Sonne (n 26) 126.

¹³⁴ Resolution on the climate emergency (n 2).

question of responsibility, asking ‘who ought to pay?’.¹³⁵ While decision-makers have the funds, livestock operators are the producers of these emissions with the ability to reduce them. This redistribution of responsibility is legally justified by reference to the Polluter Pays Principle (PPP), a key principle underlying the EU’s environmental policy,¹³⁶ which mandates that polluters (livestock producers in this case) must bear the costs of preventing, controlling and remedying their pollution. However, responsibility must be distinguished from blameworthiness,¹³⁷ demonstrated by Young’s model of social connection. The model reveals that the concept’s central aim is to change the structural processes that created a situation, by holding accountable ‘all agents who contribute by their actions to the structural processes that produce injustice’ rather than attributing guilt.¹³⁸ As responsibility arises from mere participation in the diverse institutional processes producing the climate emergency, and not through blame, farmers should receive funds to deal with the issue. For this purpose, it does not matter whether the ‘farmers’ are individuals or big corporations as in the current analysis both are blameless and deserve compensation for restructuring their activities. Overall, a sectoral declaration assists in identifying producers as participants in institutional livestock processes leading to the climate emergency, thus making them responsible for tackling it.

As the Commission is obligated to draft proposals based on the PPP, this responsibility justifies emergency laws, detailing the stakeholders who must address the climate emergency. It thus assists in determining the contents of emergency laws. By recognising livestock operators as responsible actors, emergency legislation can justifiably redirect funds and other resources to the livestock industry to assist and compel livestock farmers to shift towards more environmentally-friendly methods. Thus, a sectoral climate emergency declaration might not only create greater normative force to overcome government complacency but may further be a tool to justify the contents of emergency laws, sustaining funding and resource allocation by specifically attributing responsibility to catalysts in the sector. Farmers will arguably be convinced to change practices as increased funding combined with normative

¹³⁵ Joel Feinberg, ‘Problematic responsibility in law and morals’ (1962) 71(3) *Philosophical Review* 340, 342.

¹³⁶ Art.191(2) TFEU.

¹³⁷ Iris Young, *Justice and the Politics of Difference* (Princeton University Press 1990) 151.

¹³⁸ Iris Young, ‘Responsibility and global justice: A social connection model’ (2006) 23(1) *Social Philosophy and Policy* 102, 102-103.

force will make it apparent that adaptation is the only way forward in the current climate emergency.

4.3 A Framework to Restructure the EU Livestock Sector

To conclude, this paper recommends an improved framework for regulating livestock emissions. Exploiting the value of a sectoral declaration, it is proposed that such a framework must be facilitated under emergency legislation, allowing the Commission greater discretion over the EU budget and mitigation policy in this area. It would thus address the institutional balance of power issue previously identified by primarily awarding decision-making powers to supranational actors such as the Commission. To recognise the temporality of emergency law, such an emergency delegation would be limited in terms of time. The framework is reliant on an economic incentives/market-based approach implemented through existing mechanisms, the CAP and ETS, as opposed to a regulatory approach.¹³⁹ Such an approach introduces flexibility and focuses on result-based reduction and reducing monitoring costs, which are crucial for the livestock sector. Competing concerns of food security and cultural identity need to be balanced against climate effects, and further, the monitoring of livestock emissions from non-point sources is difficult.¹⁴⁰ Since these instruments also include a revenue-raising component, those most affected by livestock emission policies should be compensated.¹⁴¹

i. PILLAR 1: Rewarding Climate-Friendly Practices

Subsidies: Climate subsidies are a form of financial government support for rewarding climate-friendly activities.¹⁴² In a revised framework for livestock emission regulation, abatement subsidies where farmers receive a

¹³⁹ Godefroy Grosjean and others, 'Options to overcome the barriers to pricing European agricultural emissions' (2016) 18(2) Climate Policy 151, 155.

¹⁴⁰ 'Economic Incentives' (EPA, 2022)

<<https://www.epa.gov/environmental-economics/economic-incentives#:~:text=Subsidies%20are%20forms%20of%20financial,tax%20treatment%2C%20and%20procurement%20mandates>> accessed 24 March 2022; James Breen, 'Reducing Greenhouse Gas Emissions from Irish Agriculture: A market-based approach' (2012) Research in Agriculture and Applied Economics 28.

¹⁴¹ EPA (n 140).

¹⁴² *ibid.*

subsidy for reducing emissions below a predetermined cap,¹⁴³ should be introduced under the CAP cross-compliance mechanism. These would be funded by the EU and its financial institutions. This would incentivise agricultural efficiency gains, realised through improved and innovative techniques in production methods, including capturing methane from manure and improving degraded pasture and lifetime animal productivity.¹⁴⁴ A new framework must strengthen the CAP's greening mechanisms and enforce incentives to improve grazing land management in order to preserve and create carbon sinks,¹⁴⁵ enabling carbon sequestration. Ultimately, such a framework must abolish harmful subsidies, such as coupled direct payments for high-emitting outputs, to reduce livestock emissions. Although economic subsidies do not always incite the intended effects, it is argued that in combination with the emergency narrative, which pushes the idea that change is needed not only for human survival, but financial sustainability, these incentives are provided for.

ii. PILLAR 2: Punishing Environmentally Harmful Practices

Taxation: An improved framework must impose deterrent taxation mechanisms, including, for example, a Livestock Tax: a regionally homogenous tax for livestock based on their emission potential¹⁴⁶ or unit taxes based on actual measured emissions.¹⁴⁷ Taxation must also be employed to shift consumer behaviour towards climate-friendly diets directly by taxing livestock products (for example, by increasing VAT on meat and dairy products) or indirectly by shifting livestock producers' behaviour.¹⁴⁸ It is submitted that the adoption of

¹⁴³ Pierre Gerber and others, 'Policy options in addressing livestock's contribution to climate change' (2010) 4(3) *Animal* 393, 397.

¹⁴⁴ Joseph Poore, 'Reducing food's environmental impacts through producers and consumers' (2019) 360 *Science* 987, 989-990.

¹⁴⁵ David Blandford and Katharine Hassapoyannes, 'The role of agriculture in global GHG mitigation' (2018) OECD Food, Agriculture and Fisheries Papers, No.112 8/7; Jessica Bellarby, 'Livestock greenhouse gas emissions and mitigation potential in Europe' (2013) 19 *Global Change Biology* 3, 12.

¹⁴⁶ Tara Garnett, 'Climate change and agriculture: Can market governance mechanisms reduce emissions from the food system fairly and effectively?' (2012) International Institute for Environment and Development 42.

¹⁴⁷ Gerber (n 143) 396.

¹⁴⁸ Hannah Ritchie and others, 'The impact of global dietary guidelines on climate change' (2018) 49 *Global Environmental Change* 46, 53; Damian Carrington, 'EU urged to adopt meat tax to tackle climate emergency' (*The Guardian*, 2020)

such tax laws is unlikely even if a sectoral emergency declaration was imposed. Tax law is subject to a margin of appreciation and is mostly for the individual member states to implement. For example, Germany and France have immense farming lobbies which hold great power and are likely to prevent the imposition of additional taxes on meat and dairy products. However, the viability of these measures must be considered in the emergency context, which arguably imposes a threat to business-as-usual measures and may convince EU members to impose such taxes if they are convinced that this is the only way to continue production.

iii. PILLAR 3: Combined Punishing and Rewarding

ETS Inclusion: Livestock emissions must be included in the ETS as the scheme's market-based approach leads to more efficient emission reductions at a lower cost.¹⁴⁹ Farmers may harvest private net gains from participating,¹⁵⁰ thus offsetting negative effects on their income in other areas of mitigation efforts. Frameworks such as New Zealand's emissions trading scheme, which proposes the inclusion of agricultural emissions, illustrate that the inclusion of such emission sources is possible despite them comprising mostly non-CO₂ GHGs deriving from non-point sources.¹⁵¹

Ultimately, pointed emergency intervention in the livestock sector is justified. It increases emergency declarations' normative potential and overcomes EU institutional constraints which consequently allows the use of legislative emergency powers to enforce restructuring through the proposed framework. Further, it attributes responsibility, justifying the allocation of funding and resources in such a framework by recognising efforts that can structurally transform the sector.

<<https://www.theguardian.com/environment/2020/feb/04/eu-meat-tax-climate-emergency>> accessed 24 March 2022; proposed in Farm to Fork Strategy (n 63).

¹⁴⁹ Garnett (n 146) 36.

¹⁵⁰ Ignacio Dominguez and others, "Trading schemes for greenhouse gas emissions from European agriculture: A comparative analysis based on different implementation options" (2009) 90(3) *Review of Agricultural and Environmental Studies* 287, 304-305.

¹⁵¹ Gerber (n 143) 400; NZ Emissions Trading Scheme (2013)

<https://icapcarbonaction.com/en/?option=com_etsmap&task=export&format=pdf&layout=list&systems%5B%5D=48> accessed 15 December 2021. This partial mechanism is not yet enforced.

CONCLUSION

Achieving the central aims of the Paris Agreement will become increasingly difficult if non-CO₂ GHG emissions are not addressed strongly and rapidly. Climate emergency declarations are generally considered useful tools in shifting public perception from ‘future risk’ to ‘current crises’, incentivising political action from business-as-usual approaches and establishing a new status quo.

However, an investigation of the CAP and the ESR indicates that the European Parliament’s Climate Emergency Declaration fails to incite appropriate action regarding the reduction of livestock emissions. This failure might be explained by the inability of the EU institutional framework to enact appropriate emergency powers. In response, this paper proposes an emergency declaration in the livestock sector to redirect social norms toward the imposition of an improved framework for agriculture, incentivises transition to a design that stimulates climate-friendly practices and outputs, and attributes responsibility to the right parties. Such a pointed intervention is more efficient, compatible with the general EU law principle of equal treatment, and tackles issues regarding food security. Overall, a sectoral emergency declaration is an effective instrument in ending livestock emissions’ exceptional status in EU mitigation efforts and is thus crucial in the fight against climate change.

However, within the context of wider threats to the EU, it seems unlikely that such a regime will be adopted in the near future. The Union is still recovering from criticism regarding its inability to act in crisis situations, such as the COVID-19 pandemic.¹⁵² Further, the EU is increasingly polarised amid the rule of law crisis.¹⁵³ As such, it seems unlikely that the proposed framework or a similar system will be adopted anytime soon, despite the looming threat of climate change.

¹⁵² See, for example: Stefan Lehne, ‘Why Can’t Europe Cope with the Coronavirus?’ (*Carnegie Europe*, 2021) <<https://carnegieeurope.eu/strategieurope/84286>> accessed 11 March 2023.

¹⁵³ See, for example: Laurent Pech, ‘The Rule of Law as a Well-Established and Well-Defined Principle of EU Law’ (2022) 14, 107–138 <<https://link.springer.com/article/10.1007/s40803-022-00176-8>> accessed 11 March 2023.

Viability aside, further research must address mechanisms outside the emergency law perspectives to establish whether other tools are more appropriate to stimulate action in an area as complex as animal agriculture. Ultimately, livestock emissions' intense impact on global warming indicates that it should no longer be given a free pass in climate change efforts — it should instead be subject to targeted intervention.

Sovereign Therefore Limited: The Unconstitutionality of Ouster Clauses for Errors of Law under the British Constitution

Vincent Lafortune*

ABSTRACT

This paper makes a radical claim: it argues that ouster clauses that seek to prevent judicial review for errors of law are incompatible with, and thus inadmissible under, the British Constitution. In other words, the claim is that judicial review for illegality is a constitutional fundamental that not even Parliament is competent to abolish. Two arguments are deployed to support that contention. Firstly, it is argued that a full grasp of the core constitutional principle of parliamentary sovereignty logically entails Parliament's inability to oust the judicial review jurisdiction of the courts with respect to errors of law, as parliamentary sovereignty and ouster clauses are mutually exclusive concepts. Secondly, it is argued that this limitation to Parliament's legislative remit is the manifestation of a higher order law that necessarily constrains the exercise of legislative power in a democracy.

* BCL Candidate (Oxon) '23, LLB (UCL) '22. I thank Professor Rick Rawlings (UCL) for his precious feedback on earlier versions of this piece. The views expressed in this article are mine only. Address for correspondence: vincent.lafortune@wadham.ox.ac.uk.

In exceptional circumstances involving an attempt to abolish judicial review or the ordinary role of the courts, the Appellate Committee of the House of Lords or a new Supreme Court may have to consider whether this is a constitutional fundamental which even a sovereign Parliament acting at the behest of a complaisant House of Commons cannot abolish.¹

I. INTRODUCTION

A. Aim of this Article

This article contends that ouster clauses² ('ousters') that purport to oust the jurisdiction of the courts to review executive decisions for errors of law (or, interchangeably, 'illegality' or 'classic *ultra vires*') are incompatible with, and therefore inadmissible under, the Constitution of the United Kingdom.³ To be sure, the contention is not merely that ousters *should not* be enacted by Parliament because they are inimical to a healthy constitutional balance. This would be to implicitly accede to a view of the constitutional order that this paper rejects, namely that '[e]verything that happens [in the United Kingdom] is constitutional'⁴ and that where the constitutional equilibrium is to be struck is ultimately a matter for the political arena. The contention of this piece is more radical, as statutory provisions that unequivocally purport to bar the courts from submitting decisions taken by public bodies to judicial review for illegality ought to be denied legal effect by the courts, notwithstanding unambiguous intentions

¹ R (*Jackson*) v *Attorney General* [2005] UKHL 56; [2006] 1 AC 262 [302]-[303] by Lord Steyn.

² Ouster clauses are defined as '[s]tatutory provisions which seek to limit [or indeed outright oust] the ordinary jurisdiction of the court': *Anisminic Ltd v Foreign Compensation Commission* [1969] 2 AC 147 170.

³ Reference to 'review' in this paper is intended to include all the components involved in the process of judicially reviewing an administrative act for errors of law: from the scrutinising of the act's fit under the empowering primary legislation to the nullification *ab initio* of the administrative act for want of fit (whether by way of quashing order or declaration).

⁴ John A.G. Griffith, 'The Political Constitution' (1979) 42 MLR 1, 19.

to the contrary by Parliament, because such statutory provisions are unconstitutional.

This argument will be advanced in two parts. In the first part, I will argue that the core constitutional principle of parliamentary sovereignty itself logically necessitates Parliament's inability to oust judicial review for errors of law. Thus, the very thing that is often invoked as the ultimate justification for the constitutionality of ouster clauses — the notion of parliamentary sovereignty — will be shown to act as a logical bar to the abolition by Parliament of judicial review for illegality. In the second part, I will argue that this 'legislative disability', which will have been shown in the first part to be derivable from the notion of sovereignty, is in the final analysis the consequence of a higher order law that necessarily constrains the exercise of legislative power in a democracy.

B. The Type of Ouster Clauses to Be Discussed

This article deals exclusively with ouster clauses that are designed to preclude the review of executive decisions or acts falling outside the jurisdiction (in the classic *ultra vires* sense) afforded to the decision-maker by the empowering legislation.⁵ In other words, this paper tackles ouster clauses that purport to bar the courts from scrutinising and, if need be, impeaching a public body's decision for illegality. Of course, ouster clauses may also be designed to limit the applicability of the various common law principles of good administration, such as natural justice, reasonableness, proper purpose, etc. However, this paper does not address the constitutional position of these other

⁵ 'Jurisdiction' is used here in its enlarged, post-*Anisminic* sense: see *O'Reilly v Mackman* [1983] 2 AC 237, 278. Moreover, the reader may legitimately query the decision to restrict this analysis to statutorily derived public powers. Why not address public powers derived from other sources, such as prerogative powers? To echo what is said further down in the main text, this omission should not be treated as an admission. The decision to so restrict the analysis is due to multiple factors. First, statute-derived powers undeniably constitute the bulk of administrative powers in use today making their treatment of the utmost relevance. Second, the core contention is radical and controversial enough as it is that it would be unhelpful in the circumstances not to adopt as narrow a scope as possible. Third, and relatedly, the nature of prerogative executive powers is in itself a controversial topic and an analysis of ouster clauses pertaining to those powers would necessarily make assumptions that could attract criticism and thus further detract from the core case. We shall return to prerogative powers in the future.

‘precepts of legality enforced by judicial review against public agencies’.⁶ This omission should not, however, be treated as an admission that these are undeserving of the same fundamental status. It will undoubtedly be important to return to these matters in a future contribution. For now, it is simply considered that tackling ouster clauses that purport to shield administrative decisions from judicial review for errors of law is an undertaking to which one entire contribution can properly be devoted.

In principle, review for errors of law is a straightforward affair. Consider the underpinning theory to begin with, where Parliament, the legislative sovereign, grants powers to a body or a person. That body or person possesses that power only by virtue of grant of that power by the legislature; absent such a conferral, the body or person is no more legally empowered, nor no less legally burdened, than a private, lay individual. The picture that emerges is that of a primordial legal ‘black hole’, incrementally populated with ‘jurisdictional spotlights’ as Parliament empowers bodies or persons with various public administrative capacities. This imagery vividly brings to life the ‘fundamental principle, inherent throughout the legal system, that powers can be validly exercised only within their true limits’.⁷ If a power is not exercised according to its limits, the exercise of the power is but a purported one as, in truth, it is an act carried out by the body or person that falls within the legal nothingness, that is, the black hole, and however close it might have gotten to the jurisdictional spotlight, it does not benefit from its vitalising effect. The law, in other words, does not ‘see’ the act, as the ‘authority has in law done nothing’.⁸

Review for errors of law is one important way of giving effect to the above theoretical underpinnings. The court looks at the empowering statute and compares its terms with the action purportedly taken on that basis by a public body. If it turns out that the public body has acted inconsistently with the terms of the applicable law, that body has acted ‘without authority’ and its actions are in effect writ in water, having no effect at law.⁹ *UNISON*¹⁰ is instructive in that

⁶ T.R.S. Allan, *The Sovereignty of Law: Freedom, Constitution and Common Law* (OUP, 2013) 217.

⁷ William Wade and Christopher Forsyth, *Administrative Law* (11th edn, OUP 2014) Chapter 2, 11.

⁸ *ibid.*

⁹ *HM Treasury v Ahmed (No 2)* [2010] UKSC 5, [2010] 2 AC 534 [1].

¹⁰ *R (UNISON) v Lord Chancellor* [2017] UKSC 51, [2020] AC 869.

respect. In that case, the question was whether fees of access to tribunals imposed by the Lord Chancellor pursuant to Section 42(1) of the Tribunals, Courts and Enforcement Act 2007 were *ultra vires* (i.e., were mistaken as to the applicable law) because they had the effect of creating a ‘real risk that persons will effectively be prevented from having access to justice’,¹¹ an effect which the 2007 Act was interpreted as not empowering fees set under it to produce. Lord Reed found that the court fees had such an effect and that, therefore, the Order responsible for those fees was to be quashed and treated as unlawful *ab initio*.

The kind of ouster clause that this article contends is repugnant to the Constitution; it is the kind of ouster clause that seeks to prevent the courts from engaging in the process just described.

C. The Trigger for this Article

Until very recently, this paper’s relevance to public and administrative lawyers might have been legitimately doubted. That it was concerned with a question of academic interest only would have been fair criticism — although, to my mind at least, not a fatal criticism, for the quest to uncover the true nature and implications of our Constitution is surely an intrinsically worthy enterprise. The reason why the utility of this paper may have heretofore been doubted is well known to public lawyers, as ever since the case of *Anisminic* the courts have drawn the distinction between a decision and a purported decision.¹² This distinction, which was reaffirmed as recently as in 2019 in the *Privacy International* case,¹³ has effectively rendered nugatory statutory provisions seeking to preclude the courts from judicially reviewing ‘decisions’ taken by public bodies. That was so because a ‘decision’ afflicted by an error of law was in fact not a ‘decision’ but a ‘nullity’, or a ‘purported decision’, which was, as such, not captured by the ouster clause.¹⁴ In providing for the screening of ‘decisions’ from judicial review, Parliament was deemed not to have intended to preclude the review and quashing of ‘purported decisions’. Hence, there was no need to run a constitutional ‘strickdown’ argument in order to neutralise an ouster clause.

¹¹ *ibid* [87].

¹² *Anisminic* (n 2) 170.

¹³ *R (Privacy International) v Investigatory Powers Tribunal* [2019] UKSC 22; [2020] AC 491.

¹⁴ *ibid* [54].

Today, the relevance of the question at the heart of this paper — namely, whether Parliament has the competence to prevent review and quashing of executive decisions taken under an empowering statute but afflicted by one or more errors of law — cannot be in doubt. Section 2 of the Judicial Review and Courts Act 2022 (‘JRCA 2022’) is an ouster clause worded in such a way that immunises it from the linguistic distinction just discussed.¹⁵ Indeed, in providing that “‘decision” includes any purported decision”,¹⁶ Parliament closes the interpretative gap through which the courts have, until now, bypassed ouster clauses. It follows that if, and most likely when,¹⁷ s 2 JRCA 2022 is litigated, the arguments will necessarily focus on what is perhaps the most profound, contentious, and divisive of constitutional questions. Indeed, although in normal times it is certainly preferable to avoid asking whether Parliament’s legislative remit knows any limits, here the Supreme Court is bound to be asked to directly address and answer the question of Parliament’s ability to enact ouster clauses. The object of this article is therefore to suggest an answer to this critically important question.

D. Caveat

An important caveat must be borne in mind throughout this paper. As will become clear further down, one assumption upon which this contribution *in part* rests is that tribunals are not constituent entities of the judicial system. Tribunals, instead, are ‘judicialised’ executive bodies, but nothing more than that. Thus, in subscribing to the position as it was laid down by Lady Hale in *Cart*¹⁸ — where Her Ladyship clearly held the view that tribunals were not part of the ‘legal system’ (it is submitted that ‘judicial system’ is slightly more helpful terminology)¹⁹ — this article conforms to the current authoritative legal characterisation of tribunals. My view is that Lady Hale is right to say that tribunals are not judicial entities *properly so called*. Of course, one may take issue

¹⁵ This also applies to s 3 of the Dissolution and Calling of Parliament Act 2022 which predates s 2 JRCA 2022, but the ouster in Section 2 JRCA 2022 is more likely to make waves and thus is the one discussed here.

¹⁶ s 2(7) JRCA 2022.

¹⁷ This is for two main reasons. First, it is clear from the second ground of appeal in *Privacy International* (n 13) that there is an appetite to run a ‘strike down’ argument. Second, it is my hope that a flurry of academic commentary prompted by the enactment of s 2 JRCA 2022 will convince practitioners to bring such a case before the court.

¹⁸ *R (Cart) v Upper Tribunal* [2011] UKSC 28, [2012] 1 AC 663.

¹⁹ *ibid* [43].

with that assumption and with Lady Hale's characterisation of tribunals,²⁰ but it is not the aim of this paper to participate in that debate — that is a matter for another enterprise. Should the reader reject the aforementioned characterisation of tribunals, then for them the arguments in this paper will only apply to ouster clauses designed to preclude the judicial review of decisions taken by executive bodies other than tribunals.

To be clear, if one takes the view that tribunals are proper executive bodies, then the following discussion applies to ouster clauses shielding tribunal decisions as much as it does to ouster clauses shielding decisions from other, more traditional, executive bodies. But, if one takes the view that tribunals are constitutionally distinct from executive bodies and should be considered as constituents of the legal (or judicial) system, then to that person the arguments herein apply to ouster clauses attached to decisions taken by all executive bodies, but not to tribunals.

Finally, this paper should not lose its *prima facie* appeal from the standpoint of the reader who disagrees with the view that tribunals are constitutional equivalents of other executive bodies. Indeed, although for this reader the ouster clause in s 2 JRCA 2022 is immune from the arguments that will be made in this contribution because it protects a decision made by a tribunal, there are two reasons why she might still be interested in reading on. Firstly, as just mentioned, the law currently stands in favour of the view that tribunals are not judicial bodies, which means that legal arguments can confidently be run on that basis. Secondly, the government openly intends to use '[t]he explicit *Cart* ouster clause ... as an example to guide the development of effective legislation in the future'.²¹ It is likely that such future legislation will not only seek to shield decisions made by tribunals from judicial review but also decisions taken by more 'traditional' executive decision-makers. The arguments hereunder can be considered as a pre-emptive strike on such constitutional aberrations.

E. Angle of Attack

It will have become apparent to the reader acquainted with the case law on ouster clauses that this article aims to show why the second question

²⁰ Timothy Endicott, *Administrative Law* (5th edn, OUP 2021) 490.

²¹ Judicial Review Reform Consultation, 'The Government Response' (2021) CP 477 [55].

which arose but ultimately was not answered in the *Privacy International* case — namely ‘whether ... Parliament may by statute “oust” the supervisory jurisdiction of the High Court to quash the decision of an inferior court or tribunal of limited statutory jurisdiction²² — should be answered in the negative.

In the *Privacy International* litigation, it ultimately proved unnecessary for the majority²³ to answer that question. Applying the *Anisminic* interpretive approach to the construction of Section 67(8) of the Regulation of Investigatory Powers Act 2000, the majority found that Parliament had not in fact intended to oust the courts’ judicial review jurisdiction for errors of law made by the Investigatory Powers Tribunal (‘IPT’). The second question did not therefore arise for consideration. Nevertheless, some justices in the majority provided substantive *obiter dicta* on the second question, and the dissentients, having found that Parliament had in fact intended to exclude the jurisdiction of the courts to review decisions of the IPT for errors of law, necessarily had to grapple with it.

Accordingly, the approach I have chosen for this paper is to frame my arguments around Lord Sumption’s discussion of and answer to that question in his dissenting judgment.²⁴ This approach, I believe, maximises the impact of this contribution on the law. Indeed, while Lord Sumption is no longer in the Supreme Court, it is noteworthy and, in fact, of paramount importance that Lord Reed unqualifiedly agreed with Lord Sumption’s judgment in *Privacy International*.²⁵ Since Lord Reed is the acting President of the Supreme Court, His Lordship is, incontestably, an influential figure on the bench.²⁶ As things currently stand, Lord Reed’s Supreme Court would most likely follow Lord Sumption’s reasoning. Thus, instead of arguing against ouster clauses in isolation and leaving it to others to pluck and apply the substantive reasoning hereunder to the case law, I have decided to apply it myself to Lord Sumption’s judgment to show where His Lordship (respectfully) erred in concluding that Parliament did have the power to oust judicial review in that case. There are other reasons why direct engagement with Lord Sumption’s judgment is valuable, and these will become clear later.

²² *Privacy International* (n 13) [207].

²³ Lady Hale and Lords Kerr, Carnwath and Lloyd-Jones.

²⁴ *Privacy International* (n 13) [207]–[211].

²⁵ *ibid* [169].

²⁶ Conor Gearty, ‘In the Shallow End’ (2022) 42 *London Review of Books*.

Finally, as will be seen below, it is noteworthy that nothing in Lord Sumption's answer to the second question turns on the distinction between tribunals and other executive bodies.²⁷ Therefore, while the question in *Privacy International* is narrowly framed and refers only to 'inferior courts' and 'tribunals of limited statutory jurisdiction' — understandably, as the decision challenged in that case was that of the IPT — Lord Sumption's answer can safely be read as answering the broader question of whether Parliament may, by statute, oust the supervisory jurisdiction of the High Court to quash the decision of any decision-making body empowered by statute. This only further emphasises the pertinence of framing the substantive arguments of this paper around Lord Sumption's (and thus Lord Reed's) reasoning. Indeed, an attack on His Lordship's disposal of arguments against ouster clauses in a case involving tribunal decisions is, at one and the same time, a pre-emptive attack on his disposal of arguments against ouster clauses in a case involving more traditional executive decision-makers.

F. Signposting

In outline, Lord Sumption thought there were two possible arguments available to someone trying to persuade a court to deny legal effect to ouster clauses. The first, the *conceptual argument*, argued that, by denying effect to ousters, the courts would be, far from violating parliamentary sovereignty, affirming it.²⁸ Second, the *radical argument* argued that a higher law, to be 'ascertained and applied by the court[s]',²⁹ required the courts to outrightly preclude Parliament from ousting judicial review.³⁰ Ultimately, for Lord Sumption, none of these arguments proved successful, as the question of whether Parliament was competent to oust judicial review was, in the final analysis, to be answered in the affirmative.

1. *Conceptual Argument/Argument for Sovereignty*

In Section II, I argue that Lord Sumption's *conceptual argument* — which is really an *argument from sovereignty* (and which will henceforth bear that name), 'conceptual' merely reflecting the conclusion arrived at by Lord Sumption in the

²⁷ In contradistinction to Lord Wilson's answer, see *Privacy International* (n 13) [238]-[253].

²⁸ *Privacy International* (n 13) [209].

²⁹ *ibid* [208].

³⁰ *ibid*.

light of his understanding of the nature and implications of parliamentary sovereignty — is rooted in an incomplete understanding of the notion of parliamentary sovereignty. Therefore, and however unintentionally, Lord Sumption erects, and then attacks, a strawman. In truth, a fully-fledged grasp of the notion of parliamentary sovereignty leads to the conclusion that ouster clauses are logically incompatible with that notion, as parliamentary sovereignty and ouster clauses for illegality are mutually exclusive concepts. Consequently, ouster clauses cannot survive in the statute book so long as parliamentary sovereignty remains the core constitutional principle of the British legal system.³¹

This conclusion will be arrived at via a close reading, and an intertwined discussion of, Laws LJ's judgment in *Cart*³² and Lord Sumption's judgment in *Privacy International*.³³ Contrary to Lord Sumption's claim in *Privacy International* that he was merely reiterating Laws LJ's argument in *Cart*,³⁴ it will be shown that Laws LJ's version of the *argument from sovereignty* is not only fundamentally different from Lord Sumption's, but is also one which displays a true grasp of the nature, and implications, of parliamentary sovereignty. The decision to utilise and, in a sense, revitalise Laws LJ's judgment in *Cart* has important practical ramifications. With the obfuscatory veil draped by Lord Sumption over Laws LJ's argument in *Cart* lifted, the version of the *argument from sovereignty* advocated for in this article is revealed to have endorsement at the Court of Appeal level, making it a more realistic contender for the Supreme Court's seal of approval. This is another reason why direct engagement with Lord Sumption's judgment is valuable.

2. *Higher law argument*

In Section III, I argue that the fundamental problem with Lord Sumption's summary dismissal of the *radical* or *higher law argument* is that His Lordship equates the absence of a written constitution with the absence of a higher law patrolling the boundaries of the vast territory afforded to primary

³¹ A constitutional fact which was recently unqualifiedly reaffirmed in *R (Miller) v Secretary of State for Exiting the European Union* [2017] UKSC 5, [2018] AC 61 [43].

³² *R (Cart) v Upper Tribunal* [2011] QB 120.

³³ *Privacy International* (n 13).

³⁴ *ibid* [208] in conjunction with [210].

legislation.³⁵ I will endeavour to show that this is a *non sequitur*: a higher law anterior to, and immune from, primary legislation will be shown to exist and to require the striking-down of ouster clauses notwithstanding the lack of a constitutional text.

II. THE ARGUMENT FROM SOVEREIGNTY

A. Introduction

As mentioned in sub-section F above, Lord Sumption in his dissenting judgment in *Privacy International* claimed that he was reiterating Laws LJ's *argument from sovereignty* in *Cart*.³⁶ The initial aim of this section is to rebut that claim by demonstrating that the two arguments are fundamentally different. Indeed, close analysis of the relevant passages in the two judgments will reveal that Lord Sumption was not in fact reiterating Laws LJ's position, as the conclusions reached by Their Lordships with respect to the possibility of ousting judicial review are actually diametrically opposed. The upshot is that there are two versions of the *argument from sovereignty* available in the case law for the Supreme Court to choose between. Once this initial task is completed, the section will go on to demonstrate why Laws LJ's conception of parliamentary sovereignty — and therefore the Lord Justice's version of the *argument from sovereignty* — is the right one. In proceeding as such, not only is Lord Sumption's *argument from sovereignty* substantively countered, but a better alternative is revealed to have judicial endorsement.

B. The Judgments

1. *Lord Sumption*: *Privacy International*

According to Lord Sumption, refusing to give effect to an ouster clause would, 'up to a point', go towards vindicating 'the sovereignty of Parliament, and not limiting it'.³⁷ This is because it is difficult to imagine Parliament at once intending 'to create a tribunal with [both] limited jurisdiction

³⁵ *ibid* [209].

³⁶ *ibid* [208].

³⁷ *ibid* [210].

and unlimited power to determine such limit at its own will and pleasure'.³⁸ However, the contradiction in statutory terms produced by an ouster clause, far from being conclusive, was but one factor among others to be considered when interpreting an ouster clause.³⁹ The argument was, in the final analysis, 'a variant of the claimant's primary case [i.e., first ground of appeal] about Parliamentary intention'⁴⁰ and in that sense '[a] sufficiently clear and all-embracing ouster clause might demonstrate that Parliament had indeed intended to [abolish the court's judicial review jurisdiction]'.⁴¹ While it would certainly be unorthodox and 'a strange thing for Parliament to intend', it was nevertheless 'conceptually possible' for Parliament to provide as much.⁴²

2. *Lord Justice Laws: R (Cart) v Upper Tribunal*

Germane to our purposes is the passage between [32] and [39] of the Lord Justice's judgment. Laws LJ begins with the avowal of a discomfort, as His Lordship is not at ease to circumvent an ouster simply by invoking the weighty presumption, repeatedly established since *Anisminic*, against an intention of Parliament to oust the review and quash nullities. Instead, Laws LJ felt the need to explain the underlying rationale that gives rise to the presumption, lest it become nothing more than 'an article of faith'.⁴³ According to Laws LJ, the reason for the presumption is that judicial review is fundamental to the rule of law.⁴⁴ The requirements of the rule of law are many,⁴⁵ but the one that is contravened by, and which triggers the presumption against, ouster clauses is that 'statute law has to be mediated by an authoritative judicial source'⁴⁶ lest the meaning of the law 'be degraded to nothing more than a matter of opinion'.⁴⁷ Thus, 'the very effectiveness of statute law [i.e., vindication of Parliament's

³⁸ Lord Sumption in *Privacy International* (n 13) [208] citing Farwell LJ in *R v Shoreditch Assessment Committee, Ex p Morgan* [1910] 2 KB 859, 880 (emphasis added).

³⁹ *Privacy International* (n 13) [208].

⁴⁰ *ibid* [210].

⁴¹ *ibid*.

⁴² *ibid*.

⁴³ *Cart* (n 32) [33].

⁴⁴ *ibid* [34].

⁴⁵ *ibid* [35].

⁴⁶ *ibid* [36].

⁴⁷ *ibid* [38].

legislative sovereignty] ... requires' that an independent and impartial body, namely, a court, be the ultimate interpreter of statutes.⁴⁸

Having laid down the foundations atop which the robust presumption against ouster clauses rests, Laws LJ shifts gears: 'the need for ... an authoritative judicial source *cannot* be dispensed with by Parliament'.⁴⁹ These are unequivocal terms from the Lord Justice as they amount to saying that 'judicial review cannot be abolished by Parliament'.

At this point, a potential charge of internal inconsistency must be countered on behalf of Laws LJ to circumvent ambiguity and cherry-picking of his judgment. The charge is as follows: it is difficult to see how Laws LJ can simultaneously speak of a presumption against the ousting of judicial review and of a categorical impossibility for Parliament to dispense with judicial review. The nature of the presumption is that it can be displaced and yet we are told Parliament cannot even do so, as the two propositions appear mutually exclusive. I suggest, however, that these propositions can readily be reconciled. For Laws LJ, the presumption operates merely as a first line of defence against the ousting of judicial review; it is perfectly legitimate to presume both (i) that it is the legislature's intention to conduct its business in a constitutionally sound manner by keeping with the rule of law and (ii) that it is Parliament's intention for the limits it sets to be respected. In other words, the presumption is the manifestation of (i) a routine application of the principle of legality in conjunction with (ii) the same rule of interpretation as adopted by Lord Sumption. This presumption has clear operational potency as evidenced by *Anisminic* and its progeny,⁵⁰ and it is also the least contentious path to the nullification of an ouster clause. But it is not, according to Laws LJ, the ultimate justification for denying legal effect to ouster clauses, as if push came to shove and the presumption is displaced, Parliament would have to be declared outright incompetent to abolish judicial review. In other words, a presumption and a hard-edged denial of validity are not mutually exclusive, but rather two mechanisms of different intensities invoked to reach the same result, namely the nullification of ouster clauses and with it the preservation of judicial review. Whilst in most cases the presumption will be sufficient to neutralise an ouster

⁴⁸ *ibid.*

⁴⁹ *Ibid* (emphasis added).

⁵⁰ E.g., *Anisminic* (n 2), *O'Reilly* (n 5).

clause, in others it will be defeated by unequivocal language — such as an ouster clause modelled on s 2 JRCA 2022 — and recourse to hard-edged limitation of Parliament's legislative competency will be necessary, on Laws LJ's account.

C. The Fork in the Road and Challenge Posed

Despite Lord Sumption's claim that in laying down his conceptual argument he was merely reiterating 'the view suggested by Laws LJ in the Divisional Court in [*Cart*]',⁵¹ the two judgments plainly arrive at drastically different conclusions. As just seen, for Laws LJ 'the need for ... an authoritative judicial source *cannot* be dispensed with by Parliament'.⁵² In stark contrast, for Lord Sumption, Parliament may, in the ultimate analysis, dispense with judicial review for errors of law by enacting a clear enough 'all-embracing ouster'.⁵³

A choice must therefore be made between Lord Sumption's and Laws LJ's version of the *argument from sovereignty*. To inform that choice, it is necessary to appreciate that the fork in the road between the two views is the result of a fundamental disagreement as to the nature, and therefore the implications of, parliamentary sovereignty. First, for Lord Sumption, 'parliamentary sovereignty' is shorthand for 'what Parliament intends to be law is law'. This is clear from the way His Lordship reduces the conundrum posed by ouster clauses to an interpretative one. I will call this the 'thin' version of parliamentary sovereignty. Laws LJ, on the other hand, advances a substantive understanding of the notion of parliamentary sovereignty. This is clearest when two of His Lordship's claims are considered together, namely that an 'authoritative judicial source *cannot* be dispensed with by Parliament'⁵⁴ and that to limit Parliament's legislative power is in fact to 'affirm' its sovereignty.⁵⁵ Thus, for Laws LJ, the notion of parliamentary sovereignty cannot be reduced to, and thus (what comes to the same) is at least not fully captured by, the notion of giving effect to Parliament's expressed intentions without question. I will call this the 'thick' version of parliamentary sovereignty.

⁵¹ *ibid* [208].

⁵² *Cart* (n 32) [38] (emphasis added).

⁵³ *Privacy International* (n 13) [210].

⁵⁴ *Cart* (n 32) [38] (emphasis added).

⁵⁵ *ibid*.

At first glance, Laws LJ's substantive version of sovereignty seems like judicial supremacy cloaked under the pretence of affirming parliamentary sovereignty. It is true that, without more, His Lordship's assertions could appear rather arbitrary in light of the Diceyan conception of parliamentary sovereignty which was, once again, recently reaffirmed by the apex court.⁵⁶ And yet, there is one crucial clue rather casually revealed by Laws LJ which, it is argued, plants the signpost to the fundamental and true rationale underpinning the idea of a substantive conception of parliamentary sovereignty that entails Parliament's inability to enact ouster clauses. Critically, His Lordship draws a direct comparison between the 'old rule that Parliament cannot bind itself' and his rule that judicial review cannot be ousted by Parliament.⁵⁷ The implicit proposition that inheres in this direct comparison drawn by Laws LJ is the rule that Parliament cannot bind itself and shares the same rationale as the rule that Parliament cannot dispense with judicial review. Now, the fact that Parliament is disabled from binding its future self is a truism that no one, let alone Lord Sumption, would seek to deny — and indeed His Lordship acknowledges it.⁵⁸ Thus, the challenge appears straightforward: if the analogy drawn by Laws LJ is revealed to be sound, by parity of reasoning, the proponents of the 'thin' version of sovereignty have no choice but to admit the validity of Laws LJ's 'thick' version. They must also concede that however 'clear and all-embracing'⁵⁹ the ouster clause, it cannot succeed in ousting judicial review for errors of law.

D. Challenge Answered

Our inquiry into the rationale behind the rule that Parliament cannot bind itself starts with Sir William Wade's seminal article, *The Basis of Legal Sovereignty*.⁶⁰ The crucial passage for our present purposes is the following, in which Wade quotes *Salmond on Jurisprudence*:⁶¹

But whence comes the rule that Acts of Parliament have the force of law? This is legally ultimate; its source is historical only, not legal ... It is the law because it is the

⁵⁶ *R (Miller) v SS for Exiting the EU* [2017] UKSC 5; [2017] 2 WLR 583 [43].

⁵⁷ *Cart* (n 32) [38].

⁵⁸ *Privacy International* (n 13) [208].

⁵⁹ *ibid* [210].

⁶⁰ William Wade, 'The Basis of Legal Sovereignty' (1955) *Cambridge Law Journal* 172.

⁶¹ *ibid* 187.

law, and for no other reason that it is possible for the law itself to take notice of. No statute can confer this power upon Parliament, for this would be to assume and act on the very power that is to be conferred.⁶²

This is a rule of logic. Wade, developing on Salmond's observations, rightly points out that, 'if no statute can establish the rule that the courts obey Acts of Parliament[, then] no statute can alter or abolish that rule'.⁶³ Therefore, one readily notices that the fundamental reason why Parliament cannot bind its future self is that to do so would undermine parliamentary supremacy, for some later Acts would need to be struck down or ignored if they were incompatible with earlier, 'enshrined' ones. A further corollary deducible from Salmond's rule of logic must be that it cannot be possible for Parliament to divest itself of, transfer, or share its legal sovereignty — this too would be to proceed on the assumption that Parliament's legal sovereignty is a product of its own will, a proposition which Salmond has just shown is absurd as a matter of logic. Thus, in Wade's words, the 'apparent paradox that [those rules are] unalterable by [an all-powerful] Parliament turns out to be a truism' that logically flows from the ultimately political, and not legal, fact of parliamentary sovereignty.⁶⁴

The final piece of the puzzle is to demonstrate why it would be an infringement of the same rule of logic — again, that legislative sovereignty is not a product of Parliament's will — for Parliament to shield decisions afflicted by errors of law from judicial review and therefore prevent them from being nullified, whether by a quashing order or a declaration. Again, if this can somehow be demonstrated, the challenge is answered. It is helpful to picture the following scenario. An executive body acts outside jurisdiction because it has, for example, considered a matter which it was not entitled to consider on its way to reaching a determination.⁶⁵ Now, realise that for as long as its decision stands, it has in essence both crafted its own jurisdiction — or, what comes to the same, *legislated* it — and acted pursuant to said *sui generis* legislation. In the usual course of things, of course, the public body's decision would be quashed or declared unlawful *ab initio* by a court for want of a legal basis. Yet, *ex hypothesi*,

⁶² *ibid*; Sir John Salmond, *Salmond on jurisprudence* (10th edn, Sweet & Maxwell 1947) 155 (emphasis added).

⁶³ Wade (n 60) 187.

⁶⁴ *ibid* 188.

⁶⁵ These are essentially the facts of *Anisminic* (n 2).

this otherwise impeachable act is now rendered unimpeachable by operation of an ouster clause enacted by Parliament which precludes the court from reviewing a flawed decision. What has just occurred? Effectively, Parliament has vested the notional public body with *de facto* legislative competence; the executive has been allowed to ‘write [its] own laws’.⁶⁶ And in granting another body legislative competence, Parliament has done the very thing which the rule of logic expounded above has shown it is logically disabled from doing — because, once again, it is not the source of its own legislative supremacy — namely, transferring or sharing its legislative sovereignty with another body.

The challenge has therefore been answered. Ouster clauses are plainly incompatible with an orthodox Diceyan conception of parliamentary sovereignty. What is often invoked as the ultimate justification for Parliament’s power to enact ouster clauses if it so wishes, namely, the core constitutional principle of parliamentary sovereignty, has been shown to disable Parliament from doing that very thing.

E. Briefly Addressing Criticism

Some commentators have expressed scepticism at Laws LJ’s analogy between the old and trite rule that Parliament cannot bind itself, and His Lordship’s strict rule that Parliament cannot oust judicial review. In particular, Aileen McHarg objects that Laws LJ’s analogy is ‘far from compelling’ and a tad deceiving.⁶⁷ That is so because, in her view, while the former rule is derived from the fact that every iteration that Parliament enjoys ‘unlimited legislative competence’, the latter rule is one which stems from the sacrosanctity of the ‘rule of law’.⁶⁸ Therefore, to McHarg, while the former rule is a product of the core constitutional notion of parliamentary sovereignty, to suggest as much of the latter rule by drawing an analogy with the former is to ‘turn the doctrine [of parliamentary sovereignty] on its head’.⁶⁹

⁶⁶ *Cart* (n 32) [37].

⁶⁷ Aileen McHarg, ‘Giving Substance to Sovereignty: Parliamentary Sovereignty and Parliamentary Effectiveness’ in Brice Dickson and Conor McCormick (eds), *The Judicial Mind: A Festschrift for Lord Kerr of Tonaghmore* (Bloomsbury 2021) 211.

⁶⁸ *ibid.*

⁶⁹ *ibid* 212.

In this section, I have attempted to show that the analogy is actually sound and that both rules share the same source. In other words, McHarg's 'distinction of sources' argument has been shown to be unfounded. Indeed, both the rule that says Parliament cannot bind its future self and the one that says Parliament cannot dispense with judicial review for errors of law are corollaries of, and thus are derived from, the notion of parliamentary sovereignty. Ouster clauses are not denied effect because 'they are deemed to be incompatible with the rule of law'.⁷⁰ They are denied effect because they run counter to parliamentary sovereignty when that notion is properly understood. As shown, once it is realised that the rule that courts continuously obey Acts of Parliament (i.e., parliamentary sovereignty) is of extra-legal nature (i.e., that it is not a rule that emanates from Parliament itself), it becomes clear that Parliament cannot meddle with that rule, which it would be doing if it purported to alter its own legislative competence (which gives us the rule against entrenchment of legislation), or to transfer or share its legislative authority (which gives us the rule against ouster clauses for errors of law). In other words, Parliament's legislative authority must remain both limitless *and* exclusive. And since an ouster clause for errors of law is tantamount to Parliament transferring its legislative authority to another body, denying it legal effect is not to violate but to preserve Parliament's *exclusive* legislative authority.

F. Conclusion

There are two key takeaways from this section. Firstly, the veil cast over Laws LJ's argument by Lord Sumption has been lifted. It is simply not the case that for Laws LJ too the dilemma is merely 'conceptual' and interpretative. There are two starkly contrasting versions of the *argument from sovereignty* to be found in the case law. This is important, as the Supreme Court has a choice between two alternatives that have curial support. Secondly, this article has shown that the correct choice is Laws LJ's strict version of the argument since it flows from a 'thick', and true, conception of sovereignty. Indeed, a careful examination of the notion of parliamentary sovereignty reveals inbuilt disabilities derived from the fact that Parliament is not the source of its own legislative supremacy. One of the said disabilities is that Parliament cannot bind its future self. Another is that it cannot transfer, or share, its legislative competence. And since ouster clauses for errors of law result in Parliament

⁷⁰ McHarg (n 67) 211.

effectively granting executive legislative competence, such ouster clauses must be denied legal effect so long as parliamentary sovereignty continues to reign supreme. In the name of parliamentary sovereignty, ouster clauses must not be given legal effect in this legal system.

III. THE HIGHER LAW ARGUMENT

A. Introduction

According to Lord Sumption, the radical argument contends that courts may, in exceptional circumstances, strike down statute in the name of a 'higher law'.⁷¹ His Lordship rejects that argument on the basis that: '*in the absence of a written constitution* capable of serving as a higher source of law', primary legislation is supreme.⁷² The purpose of this section is to show that this is a *non sequitur*: the lack of a textual constitution does not entail a completely limitless legislative remit for Parliament. Written constitutions recognise, rather than create, principles and rights fundamental to, and ever-present in, democratic societies.⁷³ Accordingly, these principles and fundamental rights are also present in democratic societies with unwritten constitutions, such as the Constitution of the United Kingdom.

The challenge this section sets out to meet can be formulated as follows. If it can be shown, as it will be, that in attempting to preclude the quashing of executive decisions or acts that are afflicted by an error of law, Parliament is seeking to use its legislative powers in a way that runs counter to the justification for its very own legislative empowerment, then it must follow that it is disabled from doing so. Importantly, the contention here is not merely that 'we *should not* attribute to Parliament intentions that flout the very conditions on which the legitimacy of its enactments depends'.⁷⁴ Rather, it is that notwithstanding such intentions, eminently clear as they may be, they cannot prevail.

⁷¹ *Privacy International* (n 13) [208].

⁷² *ibid* [209] (emphasis added).

⁷³ *R (Daly) v Secretary of State for the Home Department* [2001] UKHL 26; [2001] 2 AC 532, 548.

⁷⁴ Allan (n 6) 219 (emphasis added).

The argument made in this section buttresses the *argument from sovereignty* made in the previous section. It serves to show that the notion of parliamentary sovereignty and its logical ramifications — one of which being that Parliament may not oust the judicial review jurisdiction of the courts to quash decisions afflicted by errors of law — far from being products of sheer happenstance, are manifestations of the necessary limitations on legislative power that operate in democratic political systems.

B. Democracy and its Underpinning Ideals

The starting point, and one fundamental assumption for the purpose of this section, is that the United Kingdom is a democracy. This is not a contentious assumption. It is supported by the observable fact that the lawmaker is Parliament, an entity constituted of representatives elected from time to time via universal suffrage. In Britain, therefore, the wielding of political power is contingent on the continuing assent of the people, as '[t]he electorate can throw out the government'.⁷⁵ That is the cornerstone feature of the democratic form of government. It follows that it is the ideals that underpin democracy which need unearthing to assess whether, in light of 'the constraints inherent in the best conception of democracy',⁷⁶ ouster clauses are admissible under the British Constitution.

It is helpful, in seeking to identify democracy's underpinning ideals, to take as a starting point its antithesis, autocracy.⁷⁷ To understand the motivation behind the shift from autocracy to democracy — and therefore to grasp the ideals that underpin, and that are sought to be given effect to by, democratic forms of government — we must understand, at least in outline, the lessons often attributed to the Age of Enlightenment. Distilled to its essence, one of the main takeaways from the philosophical revolution of the 17th and 18th centuries was that no person is invested with inherent moral supremacy and that any political regime that is premised on such a notion is absurd and, ultimately, wicked. Diderot, for example, made the point trenchantly when he proclaimed that 'the arbitrary government of a just and enlightened prince is always bad'.⁷⁸ To be sure, 'arbitrary government' is here used to mean the wielding of political

⁷⁵ John Laws, *The Constitutional Balance* (Bloomsbury 2015) 31.

⁷⁶ Allan (n 6) 218.

⁷⁷ Laws (n 75) 31: 'Without democracy the government is by definition autocratic'.

⁷⁸ Ritchie Robertson, *The Enlightenment* (Penguin 2020), 833.

authority justified purely on the footing of a title derived from lineage or any other factual circumstances. The vice in all such claims to authority was that normative statements ('you ought to do this') could not, without being absurd, rest on mere statements of fact ('because I am X, Y, Z.'). Indeed, as Hume famously remarked, it was 'altogether inconceivable' (and here 'inconceivable' should be understood in the sense of 'impossible') that normative statements could be deduced from factual premises, and this was precisely the rule of logic that the autocratic form of government so blatantly flouted.⁷⁹

Instead, a normative statement may only validly be rooted in an appeal to reasons (i.e., by an appeal to the goodness of the consequences which it is thought will be produced by the act). But as soon as one person claims that she has good reasons to demand that others act as she wills, the natural response, from those who disagree, is to argue that the expected consequences of the act are not in fact beneficial; or that they are but that the act suggested is not the proper way to achieve them; or that those actions should not take priority over others, etc. The ability to reason is universal and 'difference and disputation' about what ought to be done is inevitable.⁸⁰ Indeed, 'it is an inherent feature of unruly humankind that what benefits the people will always be contentious'.⁸¹ Once this is grasped, the path to democracy starts to take shape. There is, however, one last condition of the human kind (the universality of reason and the plurality of normative outlooks being the two just identified) that remains to be acknowledged.

Man is a gregarious species. No doubt there are solitary individuals, 'but society ... is the paradigm'.⁸² Indeed, 'amongst the Things peculiar to Man is his Desire of Society'.⁸³ This is an observable fact; it is hardwired in our condition as much as the ability to reason is. The upshot is that a structure to regulate collective life needs to be erected lest cacophony ensues because, as we have seen above, individuals will inevitably advance conflicting normative propositions — we need a structure to solve our unsocial sociability without falling foul of the autocratic fallacy. In other words, we need a mechanism to

⁷⁹ David Hume, *Treatise of Human Nature* (Clarendon Press, 1896), 469: 'You cannot derive an ought from an is'.

⁸⁰ Laws (n 75) 31.

⁸¹ *ibid.*

⁸² *ibid.* 7.

⁸³ Hugo Grotius, *The Rights of War and Peace* (Liberty Fund, 2005), 79.

invest political power in a ruler (whose political decisions for the community are authoritative and ultimate) which does not make a mockery of the realisations made above, namely (i) that only reasons can validly ground normative propositions and (ii) that the faculty of reason is a universal feature of the human condition.

The answer, of course, is the democratic machinery. Democracy provides a political battleground, the outcome of the tussle occurring in which is dictated by a candidate, and her normative outlook having garnered the greatest share of support from reasoning-capable human peers. That way, and for the time being, the exercise of the ruler's political authority is cured of arbitrariness in the sense that the exercise of her political authority, which manifests itself through the enactment of laws, is taken to have the tacit assent of the polity. It is only through the polling booth that 'the measures passed by government' may acquire the 'crucial moral authority' which is so blatantly lacking in autocracies.⁸⁴

C. Implications for Judicial Review

1. Generally

Having established in outline the ideals that motivate the establishment of, and therefore underpin, the democratic form of government, we are now equipped to ascertain the ground rule in accordance with which the exercise of democratically acquired political power must be conducted. That ground rule, in turn, shapes to the most minute detail the 'conditions under which [those who wield political power] are permitted to do so'.⁸⁵

The foregoing entails the following. Since it is only because it accedes to political power via a democratic procedure that the lawmaker's (Parliament's) promulgations are cured from the scourge of arbitrariness — such cure being, again, the chief aim of the democratic machinery — it must follow that *the lawmaker is disabled from empowering those taking decisions pursuant to, and acting in accordance with, its promulgations to act in an arbitrary fashion*. That is the ground rule. Had it been otherwise, the legislature would be allowing an exercise of public powers which does not advance the purpose (which is what 'to act arbitrarily'

⁸⁴ John Laws, 'Law and Democracy' [1995] Public Law 72, 85.

⁸⁵ *ibid* 79.

means in this context) of the only legitimate source of public or executive powers in a democracy, namely the legislature's own promulgations. In other words, the ruler's law would 'stifle its own justification'.⁸⁶ Thus, 'any delegation of power to a public authority must be applied to the furtherance of legitimate public purposes' which Parliament has authored.⁸⁷ It follows that Parliament does not draft its laws on a blank canvas. Instead, indelible terms and conditions invariably populate every single Act of Parliament. These provide that Parliament cannot violate the democratic justification that lends its words the power of law, which it would be doing if it permitted arbitrary acts under the law it promulgates to withstand judicial scrutiny. Law in a democracy, therefore, has a baseline substance that is 'impervious to any purported legislative abrogation';⁸⁸ law is the sovereign.⁸⁹

2. *Classic ultra vires or illegality*

The necessity of the *ultra vires* or illegality ground of review, and thus the inadmissibility of ouster clauses that prevent the quashing of executive decisions afflicted by the vice of illegality, is very straightforwardly distilled from the foregoing discussion. As this article has demonstrated earlier, a 'decision ... that flouts the limits of [the] empowering legislation' is tantamount to the executive crafting its own jurisdiction or, what comes to the same, legislating.⁹⁰ But this is pure arbitrariness: the executive does not benefit from the same democratic cure as the legislature; it cannot craft *sui generis* purposes but must give effect to Parliament's own. If Parliament were allowed to preclude the quashing of decisions plagued by errors of law, it would be actively sabotaging the democratic rationale upon which its own legislative power rests, something which, as we have seen, the ground rule necessarily prohibits.

IV. CONCLUSION

The chief aim of this article was to mount a frontal attack on ouster clauses. Particularly, this article has focused on ouster clauses designed to prevent the review and quashing of executive decisions purportedly taken

⁸⁶ Laws (n 75) 32.

⁸⁷ Allan (n 6) 224.

⁸⁸ *ibid* 217.

⁸⁹ As the title of Allan's latest book, *The Sovereignty of Law* (n 6), rightly suggests.

⁹⁰ Allan (n 6) 220.

pursuant to an empowering statute, but ultimately afflicted by one or more errors of law. With the *argument from sovereignty*, this article has shown that a fully-fledged understanding of the notion of parliamentary sovereignty, the core constitutional principle of the United Kingdom's legal system, leads to the rejection of ouster clauses. One cannot at once espouse the principle of parliamentary sovereignty and accept the validity of ouster clauses. With the *higher law argument*, I have sought to show that the foundations upon which the exercise of political power rests in a democracy entail Parliament's inability to allow decision-makers acting pursuant to its laws to act arbitrarily.

In conclusion, if this analysis has found favour with the reader, they are invited to ponder whether the above arguments can in some way be extended to capture all the grounds of judicial review currently recognised by the courts. In other words, can an argument be made, using what has been said above and especially in Section III, to say that Parliament should not be able to oust judicial review for grounds other than illegality? There is scholarship that argues for the fundamentality of all grounds of judicial review — and 'fundamentality' here should be understood to mean imperviousness to unilateral riddance by Parliament of these grounds of review.⁹¹ There is also recent high-level *obiter dicta* that points this way.⁹² No particular claims to that effect are made here, but there is no reason why we should not now direct our efforts to the inquiry of whether such a claim can be made in light of the constitutional arrangement of the United Kingdom where Parliament may make the *law* it wishes, so long of course as it does not purport to give the executive legislative competence.

⁹¹ Allan (n 6) 217.

⁹² *Privacy International* (n 13) [144].

The Role of Developing Countries in Investor-State Arbitration: Reflections on *Tethyan Copper v Islamic Republic of Pakistan*

Sarah Ahmad*

ABSTRACT

This article explores the role of developing countries in Investor-State Arbitration, using the dispute of Tethyan Copper Company Pty Limited v Islamic Republic of Pakistan as a case study. It argues that the international investment treaty regime disfavours developing countries and often poses an obstacle to sustainable economic growth and development. The article therefore proposes the construction and implementation of a good governance framework that would help rebalance the obligations between foreign investors and host States.

* Sarah Ahmad is an LLM candidate at LSE Law School, where she previously read law (LLB) as an undergraduate. Her main interests include public international law and global economic governance. The author would like to thank Oliver Hailes and the Editorial Board for their valuable comments on earlier drafts. All errors remain her own.

INTRODUCTION

When the Government of Pakistan publicised its plans to revive the controversial Reko Diq project, the announcement to resume mining activities in one of the world's largest gold and copper reserves in the Province of Balochistan was met with outrage and concern.¹ After all, Pakistan had previously lost arbitration proceedings at the International Centre for Settlement of Investment Disputes (ICSID)² as well as the International Chamber of Commerce (ICC)³ arising out of the Reko Diq project. The resulting US\$ 11 billion in awards placed a significant financial burden on an already debt-stricken country, making former Prime Minister Imran Khan call for the termination of 23 of the country's Bilateral Investment Treaties (BITs) to avoid being sued in further disputes.⁴ What prompted Pakistan, signatory to the first-ever BIT with Germany in 1959, to advocate for the suspension of today's investment treaty regime? And why did the Government decide to resume the very same project only a year later, with the Supreme Court of Pakistan constitutionally upholding the Government's US\$ 7 billion Reko Diq deal in December 2022?⁵

This article aims to explore the *Tethyan* dispute⁶ through a historical and sociopolitical lens, examining the position of developing countries in Investor-State Dispute Settlement (ISDS).⁷ It will argue that today's investment

¹ Afshan Subohi, 'Reko Diq - dealing against a tight deadline' (*Dawn*, 28 March 2022) <<https://www.dawn.com/news/1682077>> accessed 12 February 2023.

² *Tethyan Copper Company Pty Limited v Islamic Republic of Pakistan*, ICSID Case No ARB/12/1, Award (12 July 2019).

³ *Tethyan Copper Company Pty Limited v Province of Balochistan*, ICC Case No 18347/VRO/AGF, Partial Award (8 July 2019).

⁴ Manuel Pérez-Rocha, 'Ousted Pakistani Leader Was Challenging Investment Treaties That Give Corporations Excessive Power' (*Inequality.org*, 14 April 2022)

<<https://inequality.org/research/pakistan-khan-investment-treaties/>> accessed 12 February 2022.

⁵ Advisory Opinion on Presidential Reference No.2 of 2022

<https://www.supremecourt.gov.pk/downloads_judgements/reference._2_2022.pdf> accessed 12 February 2023.

⁶ *Tethyan Copper Company* ICSID case, Award (n 2).

⁷ This paper distinguishes between 'developed' and 'developing' countries to illustrate the divergent levels of development between host and home States participating in the investment treaty regime, aligned with the classification of economies by the United

treaty regime affords disproportionately extensive protections to foreign investors in comparison to the host State's legitimate public policy concerns.⁸ This is particularly onerous for developing nations, which (i) may not have the financial means to pay such high damages awards and (ii) are less likely than developed countries to maintain the high standards of transparency and security promised in investment treaties due to weaker institutional and legal structures. Part I will briefly outline the *Tethyan* saga, which serves as a point of reference in this article, contextualising the dispute and the tribunals' reasoning. Part II will provide a critical analysis of the investment treaty regime, exploring its geopolitical origins and effect on economic development. This analysis will show how ambiguous language in investment treaties allows tribunals to adopt interpretations that favour investors over developing host States. Part III will conclude with recommendations for reforming the international investment treaty regime, exploring how developing countries could participate more equally in the system and benefit from an elaborate good governance framework.

I. CONTEXTUALISING THE TETHYAN DISPUTE

This piece begins by outlining the facts of the *Tethyan* dispute in greater detail, as it is emblematic of the shortcomings of the investment treaty regime. Situated in the northern province of Balochistan, the gold and copper reservoir at Reko Diq is estimated to hold more than 5.9 billion tonnes of ore.⁹ It has attracted many foreign investors, lured by the prospect of significant economic gains. Australian mining company Tethyan Copper Company Pty. Limited (Tethyan) was a foreign investor who succeeded BHP Minerals International Exploration Inc in the existing Chagai Hills Exploration Joint Venture Agreement (CHEJVA) with the Government of Balochistan, dated 29 July 1993. From 2006 to 2011, Tethyan, later acquired by the Canadian Barrick Gold Corporation (Barrick) and Chilean Antofagasta in equal shares, invested heavily

Nation Conference on Trade and Development. No judgment is meant to be associated with the use of this terminology.

⁸ Such public policy concerns include especially health and the environment; see Rahim Moloo and Justin M Jacinto, 'Environmental and Health Regulation: Assessing Liability under Investment Treaties' (2011) *Berkeley Journal of International Law*.

⁹ *Abdul Haque Baloch v Government of Balochistan* (PLD 2013 SC 641) [2].

in the mineral exploration at Reko Diq.¹⁰ When the Government of Balochistan declined Tethyan's mining lease application on 15th November 2011, Tethyan initiated arbitration proceedings under the Pakistan-Australia Investment Treaty at the ICSID,¹¹ as well as ICC proceedings based on the CHEJVA.¹² As this article examines the role of developing countries in ISDS, subsequent analysis will disregard other commercial arbitration proceedings arising from the Reko Diq project.

The ICSID tribunal awarded damages of US\$ 5.9 billion to Tethyan.¹³ It reasoned that by refusing to extend Tethyan's lease, Pakistan had violated the standard of fair and equitable treatment.¹⁴ Pakistan was also held to have breached its obligation not to impact the claimant's investment, which de facto amounted to unlawful expropriation of Tethyan.¹⁵ This was because Tethyan had 'spent more than US\$ 240 million on its exploration work' before filing the mining lease application 'which would have allowed [the] Claimant to amortize the expenditures it had incurred during the exploration period'.¹⁶ The Government of Balochistan's decision to refuse the lease rendered useless the information gathered during this exploration period, and 'the value of both the CHEJVA and TCCP [Tethyan's local subsidiary] was effectively neutralised'.¹⁷ Moreover, the tribunal considered that the Government of Balochistan's denial of the lease application was driven by the intention to establish its own mining project and discontinue the collaboration with Tethyan, which 'excluded the classification of the denial as a *bona fide* regulatory measure'.¹⁸ However, it should be mentioned that the Government of Balochistan's refusal to extend the lease could be explained by allegations of corruption and fraud involving Government officials, deduced inter alia from the conflicts of interests by the Balochistan Development Authority's chairman.¹⁹ In fact, in *Abdul Haque Baloch*

¹⁰ *Tethyan Copper Company Pty Limited v Islamic Republic of Pakistan*, ICSID Case No ARB/12/1, Decision on Jurisdiction and Liability (10 November 2017).

¹¹ BIT Pakistan-Australia (1998) art 13(2).

¹² *Tethyan Copper Company* ICC case (n 3).

¹³ *Tethyan Copper Company* ICSID case (n 2).

¹⁴ BIT Pakistan-Australia art 3(2) and art 7(1).

¹⁵ *ibid* art 3(3).

¹⁶ *Tethyan Copper Company*, Decision on Jurisdiction and Liability (n 10) [1328].

¹⁷ *ibid*.

¹⁸ *Tethyan Copper Company* ICSID case (n 2) [157].

¹⁹ Jeffrey D Sachs, 'How World Bank Arbitrators Mugged Pakistan' (*Project Syndicate*, 26 November 2019)

vs. Government of Balochistan,²⁰ the Supreme Court of Pakistan declared void the initial CHEJVA agreement of 1993 and, consequently, the emanating agreements with Tethyan. This was because the ‘sweetheart deal’ violated the relevant provincial and federal mining legislation²¹ and, by granting BHP utmost favourable conditions to the detriment of the people of Balochistan, the benefits bestowed upon the foreign investor undermined public policy.²² This set of facts is important, as it illustrates the multifaceted nature of problems arising in the context of BITs and informs the discussion of reform proposals advanced in this piece.

On balance, the *Tethyan* case led to an arguably perverse outcome: Pakistan was held liable to pay damages equivalent to its 2019 IMF loans for a mining project that never proceeded beyond the planning stage.²³ This is reflective of wider structural, conceptual, and ideological deficiencies of the investment treaty regime. The following Part will critically analyse the present investment system, examining how economic inequality is perpetuated to the detriment of developing countries through the regime’s core elements.

II. A CRITICAL PERSPECTIVE ON ISDS: HOW THE INVESTMENT TREATY REGIME DISFAVOURS DEVELOPING COUNTRIES

Investment treaties are often hailed as ‘depoliticised’ for offering legal resolution separate from local courts.²⁴ Disputes involving foreign investors ought to be resolved peacefully in investment tribunals without instigating diplomatic tensions with the host State, while at the same time offering investors protection abroad.²⁵ After the Second World War, States intended to hedge

<<https://www.project-syndicate.org/commentary/world-bank-corrupt-arbitration-ruling-against-pakistan-by-jeffrey-d-sachs-2019-11>> accessed 12 February 2023.

²⁰ *Abdul Haque Baloch v Government of Balochistan*[122].

²¹ *ibid* [44].

²² *ibid* [44].

²³ ‘Pakistan: History of Lending Commitments as of February 29, 2020’ (*International Monetary Fund*)

<<https://www.imf.org/external/np/fin/tad/extarr2.aspx?memberKey1=760&date1key=2020-02-29>> accessed 12 February 2023.

²⁴ Ursula Kriebaum, ‘Evaluating Social Benefits and Costs of Investment Treaties: Depoliticization of Investment Disputes’ (2018) 33 *ICSID Review* 14.

²⁵ *ibid* 16.

against political risk through law. With the first Bilateral Investment Treaty (BIT) between Pakistan and Germany,²⁶ the project of economic integration through law *beyond* Europe began in 1959.²⁷ Though concluded between States, BITs include standards of protection for foreign investors in host countries. Once foreign investors are covered by a BIT, they are subject to different standards of protection than citizens in both their home and the host countries, having special legal standing under international jurisdiction.²⁸ While foreign investors are obliged to follow local laws, the treatment they are subject to will not be determined by local courts, but through judicial review by investment tribunals with the power to award damages against the host State.²⁹

2.1 The Fallacy of Reciprocity in the Investment Treaty Regime

Notably, a core feature of this regime is the doctrine of reciprocity: both host and home States equally commit to protect the respective foreign investors.³⁰ While reciprocity is established from a *legal* perspective, one-sided investment flows from developed to developing countries and economic imbalances have prevented balanced investment flows.³¹ From 1959, Germany would conclude BITs with countries where no significant investment activity would flow to Germany; Germany was therefore not exposed to investment claims because the respective host countries were at the receiving end of German investments. This imbalance of investments makes reciprocity a legal fiction rather than an economic reality. Indeed, the Office of the High Commissioner for Human Rights warned that ‘the need to attract direct foreign investment coupled with an asymmetrical access to information and negotiation capacity often leads to unequal reciprocity ... in North-South investment

²⁶ BIT Pakistan–Germany (1959).

²⁷ Kenneth J Vandeveld, ‘A Brief History of International Investment Agreements’ in Karl P Sauvant and Lisa E Sachs (eds), *The Effect of Treaties on Foreign Direct Investment: Bilateral Investment Treaties, Double Taxation Treaties, and Investment Flows* (Oxford University Press 2009).

²⁸ William W Burke-White and Andreas von Staden, ‘Investment Protection in Extraordinary Times: The Interpretation and Application of Non-Precluded Measures Provisions in Bilateral Investment Treaties’ (2007) 48 *Virginia Journal of International Law* 307.

²⁹ *ibid* 372.

³⁰ Vandeveld (n 27) 14.

³¹ *ibid* 15.

agreements'.³² By default, States that receive numerous foreign investments are at a higher risk of being exposed to claims for damages by foreign investments. In the absence of reciprocal investments by their own citizens, this risk for host States cannot be adequately offset. Thus, BITs are de facto mostly a tool of protection for (developed) home countries, while exposing (developing) host countries to excessive liability. Albeit the 'need for international cooperation for economic development, and the role of private international investment therein' is explicitly acknowledged in the preamble of the ICSID Convention,³³ the present investment treaty regime arguably fails to live up to these aspirations.

2.2 How the Core Language of BITs perpetuates Economic Imbalances and hinders Development

The imbalance between home and host States is further perpetuated by the wording of BITs. Notably, BITs adopt very ambiguous language with *open-ended* core standards of non-discrimination (maintained in national treatment and most favoured nation provisions)³⁴ as well as absolute standards of fair and equitable treatment (which protects the legitimate expectations of foreign investors from arbitrary treatment).³⁵ In the 2009 Pakistan-Germany BIT, the revised version of the 1959 treaty, both States committed to 'promote *as far as possible* investments by investors of the other Contracting State'.³⁶ Similarly, there is an obligation to accord 'investors of the other Contracting State, *fair and equitable treatment*'.³⁷ The Contracting States must also 'give *sympathetic consideration* to applications for the entry and sojourn of persons of either Contracting State who wish to enter the territory of the other Contracting

³² United Nations Human Rights Office of the High Commissioner, 'Reforming International Investment Agreements' (2018) <<https://www.ohchr.org/sites/default/files/documents/issues/development/resources/2022-08-04/Reforming-International-Investment-Agreements.pdf>> accessed 12 February 2023.

³³ Convention on the Settlement of Investment Disputes Between States and Nationals of Other States (1966) preamble.

³⁴ Okezie Chukwumerije, 'Interpreting Most-Favoured-Nation Clauses in Investment Treaty Arbitrations' (2007) 8 Journal of World Investment & Trade 597, 604.

³⁵ Nick Gallus, 'The 'fair and equitable treatment' standard and the circumstances of the host State' in Chester Brown and Kate Miles (eds), *Evolution in Investment Treaty Law and Arbitration* (Cambridge University Press 2011).

³⁶ BIT Pakistan–Germany (2009) art 2(1) (emphasis added).

³⁷ *ibid* art 2(2) (emphasis added).

State, in connection with an investment'.³⁸ From a literal reading, it is unclear what 'as far as possible', 'fair and equitable treatment' and 'sympathetic consideration' amount to and when such standards are breached. Whilst some flexibility is needed to allow tribunals to make case-by-case judgments, leaving standard clauses undefined may result in inconsistent interpretations by tribunals which tended to side with foreign investors.³⁹ This perpetuates inequalities and unequal economic gain between developed and developing countries. If host States are found to be in breach of the applicable investment treaty, they may be liable for a substantial amount of compensation.⁴⁰ This is especially concerning for developing nations, which may not have the means to pay high damages.⁴¹

Developing countries are also less likely than developed countries to live up to the high standards of transparency and security promised in investment treaties due to weaker institutional and legal structures.⁴² Indeed, Samples shows empirically that the vast majority of investors advancing claims are from high-income countries; the majority of States being sued are low or middle-income countries.⁴³ In the *Tethyan* dispute, persistent corruption allegations, leading to a deal that violated local laws, made the Supreme Court of Pakistan void the initial CHEJVA agreement.⁴⁴ Weak rule of law compliance on provincial levels allowed for a deal that largely ignored the local population's needs and their economic interests in their natural resources.⁴⁵ Illustrative of this

³⁸ *ibid* art 3(5) (emphasis added).

³⁹ International Bar Association, 'Consistency, efficiency and transparency in investment treaty arbitration' (November 2018) <https://uncitral.un.org/sites/uncitral.un.org/files/investment_treaty_report_2018_full.pdf> accessed 12 February 2023.

⁴⁰ 'Primer on International Investment Treaties and Investor-State Dispute Settlement' (*Columbia Center on Sustainable Investment*, January 2022) <<https://ccsi.columbia.edu/content/primer-international-investment-treaties-and-invest-or-state-dispute-settlement>> accessed 12 February 2023.

⁴¹ David Schneiderman, 'International Investment Law and Discipline for the Indebted' (2022) 33 *European Journal of International Law* 65, 73.

⁴² *ibid* 72-73.

⁴³ Tim Samples, 'Winning and Losing in Investor-State Dispute Settlement' (2010) 56 *American Business Law Journal* 115, 165.

⁴⁴ *Abdul Haque Baloch v Government of Balochistan* (n 9).

⁴⁵ This is a common critique of the investment treaty regime; see Nicolás M Perrone, 'The "Invisible" Local Communities: Foreign Investor Obligations, Inclusiveness, and the

is the third-party petition filed in the Supreme Court of Pakistan on 4 January 2011 on behalf of the Sanjrani tribe of the Chagai Hills.⁴⁶ The Sanjrani claimed that in the absence of the tribe's express permission, both the Government of Pakistan and the Provincial Government were not authorised to grant land rights for exploration.⁴⁷

2.3 Developing Countries' Challenges in ISDS

With the investment treaty regime arguably functioning in favour of rich countries and foreign investors, one may question why Pakistan, after losing the *Tethyan* dispute, chose to enter into a settlement agreement and resume exploration activities with foreign investors at the Reko Diq mine. Historically, as testimonies from witnesses show, developing countries vastly underestimated the obligations they were about to enter into.⁴⁸ BITs were often negotiated by non-lawyers who did not fully understand the nature of the investment treaty regime and developing countries intended to show good cooperation by signing them.⁴⁹ Poulsen argues that '[t]he majority of developing countries thereby signed up to one of the most potent international legal regimes underwriting economic globalization without even realizing it at the time'.⁵⁰ As a former Chilean negotiator conceded, 'we signed a lot of treaties not knowing sometimes what we were committing ourselves to'.⁵¹ In Pakistan's case, when signing the very first BIT in 1959, no one could have predicted how the investment treaty regime would develop and how arbitral tribunals would interpret BITs.⁵²

International Investment Regime' (2019) 113 *American Journal of International Law* Unbound 16.

⁴⁶ Constitution Petition No.1 of 2011, Qazi Siraj-ud-Din Sanjrani and another Versus Federation of Pakistan & others. See also Oliver Hailes, 'Tethyan Copper Company Pty Limited v. Islamic Republic of Pakistan' (2022) 20 *ICSID Reports* 453, 463.

⁴⁷ *ibid.*

⁴⁸ Lauge N Skovgaard Poulsen, *Bounded Rationality and Economic Diplomacy: The Politics of Investment Treaties in Developing Countries* (Oxford University Press 2015) 109.

⁴⁹ *ibid.*

⁵⁰ *ibid.* xvi.

⁵¹ Lauge N Skovgaard Poulsen, 'Bounded Rationality and the Diffusion of Modern Investment Treaties' (2014) 58 *International Studies Quarterly* 1, 10.

⁵² The exchange of notes between Pakistan's Ministry of Finance and Germany's Ministry of Foreign Affairs shows that both parties, but especially Pakistan, attempted to narrow the scope of provisions after signing the 1959 BIT. This is illustrative of the overall tentativeness and unknowingness about how investment treaties would come to work. See 'Pakistan and Federal Republic of Germany Treaty for the Promotion and Protection

Consequently, the Pakistani Government could not foresee the complications of vague treaty language or excessively high damages.⁵³ Strikingly, when the country was first sued by Switzerland in 2001, the Government had to ask Switzerland for the final text of the treaty because it could not find a copy.⁵⁴ This shows how, most likely, the potential economic harm caused by international investment treaties was grossly underestimated and not taken seriously enough by Pakistan and many other countries.⁵⁵

Since 1959, a very close-knit community of arbitrators, who have tended to interpret generic provisions in BITs in a very investor-favoured manner, has emerged.⁵⁶ As Bjorklund and others argue, a ‘chronic diversity deficit’⁵⁷ detrimentally affects the legitimacy of ISDS: arbitrators are ‘repeat players’ who may be ‘be incentivized to have, or at least be perceived as having, a tendency to decide in favor of those ... [making their] appointments’.⁵⁸ In the *Tethyan* case, a member of the arbitral tribunal used the ‘same expert put forward by TCC [Tethyan] for another case in which the arbitrator is acting as counsel’.⁵⁹ Despite said arbitrator’s failure to disclose this conflict of interest, the other two arbitrators rejected Pakistan’s demands to disqualify him and proceedings continued as usual.

of Investments (with Protocol and exchange of notes). Signed at Bonn, on 25 November 1959’

<<https://treaties.un.org/doc/Publication/UNTS/Volume%20457/volume-457-I-6575-English.pdf>> accessed 12 February 2023.

⁵³ *ibid.*

⁵⁴ Pia Eberhardt, ‘The Zombie ISDS’ (*Corporate Europe Observatory*, March 2016)

<https://corporateeurope.org/sites/default/files/attachments/the_zombie_isds_0.pdf> accessed 12 February 2023.

⁵⁵ Poulsen (n 48) 12.

⁵⁶ Andrea K Bjorklund and others, ‘The Diversity Deficit in International Investment Arbitration’ (2020) 21 *Journal of World Investment & Trade* 410.

⁵⁷ *ibid* 440.

⁵⁸ Steven Ratner and others, ‘Lack of Independence and Impartiality of Arbitrators’ (2019) ISDS Academic Forum Working Group 6

<https://www.cids.ch/images/Documents/Academic-Forum/6_Independence_-_WG6.pdf> accessed 12 February 2023.

⁵⁹ Sachs (n 19). See also Ina C Popova and Jessica L Polebaum, ‘Emerging Expectations for Arbitrators: “Issue Conflict” in Investor-State Arbitration and Beyond’ (2018) 41 *Fordham International Law Journal* 937, 942.

ICSID proceedings are a very expensive undertaking for host States, with the arbitration costs for the *Tethyan* award in 2019 exceeding US\$ 3.7 million in total.⁶⁰ Even more concerning is the trend of multi-billion-dollar awards that have the potential to seriously disrupt economic stability in developing countries.⁶¹ In the *Tethyan* case, Pakistan argued that ‘immediate enforcement would entail immediate payment of the full amount of the US\$ 5.9 billion award that would have an “immediate and potentially devastating effect on Pakistan’s fragile economy”’. It would cause the ‘removal of funding for health, social, and welfare programs’ that would have ‘disastrous impacts for the people of Pakistan ... particularly the most disadvantaged and vulnerable’.⁶² For illustration, the ‘atypical[ly]’⁶³ high US\$ 5.9 billion award amounted to nearly as much as the IMF loan given to the country in 2019 and made up approximately 10% of Pakistan’s total annual budget.⁶⁴ Immense economic pressure, and the need to avoid fighting enforcement proceedings of the arbitral award in multiple jurisdictions,⁶⁵ thus explain the revival of the Reko Diq project. In light of this, it is questionable how the ICSID, the investment tribunal of the World Bank, can claim to be genuinely committed to facilitating development while ‘mugg[ing]’⁶⁶ developing countries through crippling ‘[b]illion-dollar-plus awards’ that fundamentally threaten their economic order.

Yet, despite the controversy around the investment treaty regime and its evident dysfunctions, developing countries are hugely dependent on foreign investors’ contributions to their economies. The *Tethyan* case clearly illustrates this. After the 2019 ICSID award, then Prime Minister Imran Khan announced

⁶⁰ *Tethyan Copper Company* ICSID case (n 2) [1847].

⁶¹ Martins Paparinskis, ‘Crippling Compensation in the International Law Commission and Investor–State Arbitration’ (2022) 37 ICSID Review 289.

⁶² *Tethyan Copper Company Pty Limited v Pakistan*, ICSID Case No ARB/12/1, Decision on Stay of Enforcement of the Award (17 September 2020) para 143. See *ibid* 290.

⁶³ Schneiderman (n 41) 67.

⁶⁴ The country’s budget for 2019–20 was Rs 8,238.1 billion (approximately US\$ 54.40 billion); see Ministry of Finance, ‘Budget in Brief’ (11 June 2019) <https://www.finance.gov.pk/budget/Budget_in_Brief_2019_20.pdf> accessed 12 February 2023.

⁶⁵ See *Tethyan Copper Company Pty v Islamic Republic of Pakistan* 590 F. Supp. 3d 262 (D.D.C. 2022). See also *Tethyan Copper Company Pty v. Islamic Republic of Pakistan* BVIHC (COM) 2020/0196.

⁶⁶ Sachs (n 19).

plans to suspend the country's BITs.⁶⁷ However, this threat has never been economically viable. On the one hand, the Government needed to enter into a settlement agreement with Tethyan Copper to resume the Reko Diq project in order to offset the immense damages it was held liable for. Without this settlement, Pakistan's strained relationship with the Financial Action Task Force (FATF) would have been further exacerbated and IMF loans jeopardised: Pakistan was put on the FATF's grey list in 2018 due to its 'strategic counter terrorist financing-related deficiencies'.⁶⁸ Compliance with FATF guidelines was one of the strict conditions to receiving the US\$ 6 billion IMF loans.⁶⁹ On the other hand, given the economically precarious situation in the country at the time and the looming threat of default, foreign investments were urgently needed to prevent the economy from collapsing. With the reconstitution of the Reko Diq project comes the creation of employment opportunities and valuable investments into local, underdeveloped infrastructure, which generates much-needed economic activity.

Thus, developing countries are trapped in a system of dependency on foreign investments. Chimni criticised that 'inequitable international investment laws' perpetuate a 'colonial vision of development for the global south that is all about "development of resources not of people"'.⁷⁰ Similarly, Gathii and Puig argue that international investment law offers no answers to the displacement of local communities, environment pollution, and destruction.⁷¹ Such arguments align closely with demands for a New International Economic Order, which call

⁶⁷ Subohi (n 1).

⁶⁸ 'There and back again: A timeline of Pakistan's journey out of the FATF 'grey list' (*Dawn*, 21 October 2022) <<https://www.dawn.com/news/1694958>> accessed 12 February 2023. See also 'Pakistan's progress in strengthening measures to tackle money laundering and terrorist financing' (*Financial Action Task Force*, 2022) <<https://www.fatf-gafi.org/en/publications/mutualevaluations/documents/fur-pakistan-2022.html>> accessed 12 February 2023.

⁶⁹ *Dawn* (n 69). See also Atlantic Council, 'Implications of Pakistan's removal from the FATF greylist' <<https://www.atlanticcouncil.org/event/implications-of-pakistans-removal-from-the-fatf-greylist/>> accessed 12 February 2023.

⁷⁰ Bhupinder S Chimni, 'Capitalism, Imperialism, and International Law in the Twenty-First Century' (2012) 14 *Oregon Review of International Law* 17, 30.

⁷¹ James Gathii and Sergio Puig, 'Introduction to the Symposium on Investor Responsibility: The Next Frontier in International Investment Law' (2019) 113 *American Journal of International Law* 1.

for the ‘restructuring of the global legal and economic order’.⁷² These demands have been largely ignored by developed countries.⁷³ As Salomon argues, industrialised countries ‘were unwilling to enter into genuine global negotiations; and then, as now, were unwilling to see rules created that did not serve their economic interests well’.⁷⁴ The relative failure of the New International Economic Order contributed to an investment treaty regime that urges developing countries to prematurely liberalise their economies and offer investment opportunities to the benefit of foreign traders.⁷⁵ Thus, the investment treaty regime’s ability to impede economic development and perpetuate economic dependencies remains a concern to this date. As Part III will argue, reform proposals need to address the foregoing critique points and consider developing countries’ challenges faced in their role as host States more holistically.

III. REFORMING ISDS: BUILDING A MORE EQUITABLE INVESTMENT TREATY REGIME

With its downsides for developing countries and pertinent transparency and fairness concerns, the ISDS regime has attracted a lot of controversy.⁷⁶ After the *Tethyan* award was rendered, justifiable complaints emerged that ‘global economic governance is broken’.⁷⁷ A fairer investment treaty regime needs to rebalance the obligations of investors and host States. This requires elaborate good governance standards for both foreign investors and host States, developed through consultation of all stakeholders.

⁷² See Charter of Economic Rights and Duties of States 1974 (General Assembly Resolution 3281 XXIX); Gleider Hernández, *International Law* (Oxford University Press 2019) 563.

⁷³ Hernández (n 73).

⁷⁴ Margot E Salomon, ‘From NIEO to Now and the Unfinishable Story of Economic Justice’ (2013) 62 *International and Comparative Law Quarterly* 31, 46.

⁷⁵ *ibid.*

⁷⁶ Jane Kelsey and Lori Wallach, “Investor-State” Disputes in Trade Pacts Threaten Fundamental Principles of National Judicial Systems (*Public Citizen*, April 2012) <<https://www.citizen.org/wp-content/uploads/isds-domestic-legal-process-background-brief.pdf>> accessed 12 February 2023.

⁷⁷ Kyla Tienhaara, ‘World Bank ruling against Pakistan shows global economic governance is broken’ (*The Conversation*, 22 July 2019) <<https://theconversation.com/world-bank-ruling-against-pakistan-shows-global-economic-governance-is-broken-120414>> accessed 12 February 2023.

3.1 Designing an Effective Good Governance Framework

This paper, therefore, proposes the implementation of a good governance framework, a set of ‘good conduct’ rules that lay the foundation of negotiations between foreign investors and host States. Such a good governance framework would provide a set of ‘best practice’ standards, outlining the expected responsibilities and legal duties of host States as well as foreign investors. Joint venture agreements should incorporate these standards, which may be industry-specific and should be phrased in plain language to ensure that all actors are cognizant of their obligations. These guidelines would inform future negotiations of joint venture agreements and ought to contribute to greater transparency in the investment treaty system. The proposed governance framework should also provide a set of model provisions to be incorporated into joint venture agreements between host States and foreign investors. Such rules should outline the respective responsibilities extensively and in detail, to counter the vague language in BITs. This may reduce the likelihood of escalating disputes that need to be referred to ISDS: the clearer the joint venture agreement and the obligations of both the foreign investor and the host State, the less the need to rely on opaque and ambiguous BITs. The proposed good governance framework thus aims to prevent the escalation of disputes and to reduce the reliance on ISDS.

The framework needs to consider more holistically the challenges host States face and must adopt a clear stance on foreign investor misconduct. As Sattorova argues, ‘[i]f investment treaties and arbitral tribunals turn a blind eye to illegal acts perpetrated by foreign investors in host states, including bribery and corruption, the investment treaty regime could be complicit in encouraging and perpetuating inadequate and undesirable patterns of behaviour’.⁷⁸ In the *Tethyan* dispute, allegations of corruption of local provincial government officials, which seemingly led to an inequitable and wholly unfavourable joint venture agreement, led the Supreme Court of Pakistan to void the CHEJVA

⁷⁸ Mavluda Sattorova, ‘Do Developing Countries Really Benefit from Investment Treaties? The impact of international investment law on national governance’ (*Investment Treaty News*, 21 December 2018) <<https://www.iisd.org/itn/en/2018/12/21/do-developing-countries-really-benefit-from-investment-treaties-the-impact-of-international-investment-law-on-national-governance-mavluda-sattorova/>> accessed 12 February 2023.

(though the arbitral tribunal's strict standards of proof for corruption were ultimately not met).⁷⁹ The proposed good governance framework would aim to discourage such behaviour by imposing more stringent transparency requirements on foreign investors and representatives of host States. This could include more stringent provisions requiring the disclosure of potential conflicts of interest in the case of an ISDS dispute. It is hoped that such measures, aiming to reduce conflicts of interests and consequently the risk of corruption, also help strengthen the rule of law compliance on provincial levels.

Good governance rules should also contain a set of recommendations designed to ensure that the host State's local populations benefit from foreign investors' activities. Such recommendations could pertain to more equitable ownership structures of investment projects, set levels of sustainable investment into local infrastructure, or the creation of jobs. Pursuant to the CHEJVA agreement, Tethyan obtained 75% of the shareholding and the Provincial Government of Balochistan only the remaining 25%.⁸⁰ As the Supreme Court of Pakistan voided this initial agreement, the reconstituted Reko Diq deal of December 2022 achieved more diversified ownership structures, with Barrick owning 50% of the project.⁸¹ In the revived project, Barrick also committed to establishing community development projects, rather than merely exporting unprocessed copper.⁸² Such measures help empower local communities and contribute to greater mutual benefit of the investment activity.

In this context, special attention to environmental protection should be paid. As with the Reko Diq project, many foreign investment projects potentially harm the environment through mining operations or other activities that extract natural resources.⁸³ A good governance framework must, therefore, strive to safeguard the local environment and mitigate climate change, thereby mitigating investment projects' possibly detrimental environmental impact. This

⁷⁹ Hailes (n 46) 469; *Abdul Haque Baloch v Government of Balochistan* (n 9) [84].

⁸⁰ *ibid.*

⁸¹ 'Reko Diq Pakistan' (*Barrick*)

<<https://www.barrick.com/English/operations/reko-diq/default.aspx>> accessed 12 February 2023.

⁸² 'Barrick Targets 2028 for First Production from Reko Diq' (*Barrick*, 16 January 2023)

<<https://www.barrick.com/English/news/news-details/2023/barrick-targets-2028-for-first-production-from-reko-diq/default.aspx>> accessed 12 February 2023.

⁸³ Jason Fry KC and Louis-Alexis Bret, 'Introduction' in Jason Fry KC and Louis-Alexis Bret (eds), *The Guide to Mining Arbitrations* (2nd edn, Global Arbitration Review 2021).

could include advancing new standard provisions in joint venture agreements which would give host States sufficient leeway for domestic policy-making to build climate resilience. Importantly, developing countries are more vulnerable to the effects of climate change due to the ‘economic importance of climate-sensitive sectors ... and their [comparatively] limited human, institutional, and financial capacity to anticipate and respond to the direct and indirect effects of climate change’.⁸⁴ In fact, climate disasters exacerbate existing economic crises by destroying infrastructure, impacting human health and lowering countries’ productivity levels.⁸⁵ This can be seen from the devastating effects of floods in Pakistan which caused over US\$ 40 billion of damage and further deteriorated levels of poverty.⁸⁶ The proposed good governance framework, cognizant of the link between climate change and development, thus needs to advocate for the inclusion of environmental protection standards in joint venture agreements.

In this regard, establishing a comprehensive *institutional* framework for ISDS could help find a more adequate balance between the protection of foreign investors and host countries’ public policy-making capabilities. Such an institution could render authoritative guidelines for interpreting existing BITs and create rules for the drafting of future treaties. It could replicate the good governance guidelines hitherto proposed and make them binding, adding enforcement power to the system. Whilst a multilateral institution could exert the necessary pressure on its members to drive change and strengthen host States’ ability to act against foreign investors, the creation of such an institution

⁸⁴ Organisation for Economic Cooperation and Development, ‘Poverty and Climate Change: Reducing the Vulnerability of the Poor through Adaptation’

<<https://www.oecd.org/env/cc/2502872.pdf>> accessed 12 February 2023.

⁸⁵ Renee Cho, ‘How Climate Change Impacts the Economy’ (*Columbia Climate School*, 20 June 2019)

<<https://news.climate.columbia.edu/2019/06/20/climate-change-economy-impacts/>> accessed 12 February 2023.

⁸⁶ Brian Bushard, ‘Record Flooding: \$40 Billion Of Damage In Pakistan As Monsoons Devastate South Asia’ (Forbes, 19 October 2022)

<<https://www.forbes.com/sites/brianbushard/2022/10/19/record-flooding-40-billion-of-damage-in-pakistan-as-monsoons-devastate-south-asia/>> accessed 12 February 2023.

Shiraz Ali Shah and Khusrav Sharifov, ‘Pakistan floods emphasize the need for tackling development from a different perspective’ (UNDP, 7 February 2023)

<<https://www.undp.org/blog/pakistan-floods-emphasize-need-tackling-development-different-perspective>> accessed 12 February 2023.

presupposes political willingness by enough States. The absence of such, however, has rendered the creation of a single comprehensive institution unfeasible.⁸⁷ The development of a good governance framework is therefore instrumental to achieving a more balanced investment treaty regime. It is hoped that the proposed standards, creating better-drafted and more equitable joint venture agreements that adequately protect both foreign investors' and host States' interests, reduce the number of disputes to be decided in the ISDS system.

CONCLUSION

This paper has aimed to critically analyse the current investment treaty regime from a historical and socio-political lens, adopting the perspective of a developing country. Using the *Tethyan* dispute as a case study emblematic of broader dysfunctions within the investment treaty regime, this contribution has sought to illustrate how the notoriously vague provisions and standards in investment treaties, combined with broader structural economic issues (like one-sided investment flows) and transparency issues in ISDS, perpetuate economic imbalances between developed and developing countries. To build a more equitable system that allows developing countries to prosper, this paper has proposed the construction and implementation of a good governance framework. Such a framework ought to inform negotiations of joint venture agreements and help rebalance obligations between foreign investors and host States. Consequently, it should be tailored to the specific challenges foreign investors and host States face, taking into consideration the development dynamics embedded and perpetuated in the system. Developing countries should form part of the investment treaty regime because they derive genuine economic benefit from it, not because they are economically dependent on maximising foreign investments and fear geopolitical repercussions.

⁸⁷ Since the rejection of the 'Singapore Issues', which included a working group for investment, during the World Trade Organization's Ministerial Conference in 1996, no multilateral investment institution could be established. See Peter Muchlinski, 'The Rise and Fall of the Multilateral Agreement on Investment: Where Now' (2000) 34 *International Lawyer* 1033.



© **LSE Law Review Volume 8, 2023**

Cover Design by Amit Pandya

Cover Photo by Nigel Stead