ISSN: 2753-5746

CAMBRIDGE LAW REVIEW

VOLUME 8
ISSUE 1
SPRING 2023

Cambridge Law Review
Editorial Board

Editors

Sebastian Kjærnli Aguirre
Ollie Bacon
Neelabh Kumar Bist
Rupal Chhaya
Marcus Choy
Thajaswini Coimbatore Balasubramanian
Cameron Coleman
Benjamin Conlon
Renatus O F Derler
Spyridon Drakopoulos
Wednesday Eden
Zhi Geng
Adhiraj Gupta
Ryan Henry Ho
Eponine Howarth
Sadie Hughes
Niharika Julka
Mehera Kochar
Varad Sumant Kolhe
Kui Li
Harry Maunder
Su Jia Moh
Leo Pang
Kamakshi Puri
Matthew Raffety
Aparna Raju
Dikshit Sarma Bhagabhati
Srijan Somal
Ishita Soni
Devina Srivastava
Christopher Symes
Elle Wintle
Kai Yeo
Cameron Zhang

International Editors

Jocelyn Catenacci
Western University
Alan Gül
Leiden University
Amanda Huang
University of Virginia
Nicholas Hui
University of Melbourne
Miranda Lalla
Leiden University
Blair Michael Ntambi
University of Cape Town
Lena Raxter
University of Ottawa; American University
Rosa Maria Torraco
University of Bologna

EDITORIAL

It is with great pleasure that we present the Spring Issue of Volume 8 of the *Cambridge Law Review*. Now in its eighth year, the journal has continued to receive many remarkable submissions from around the world, a testament to its profile and reputation, and—more importantly—to the exceptional quality of the work of previous editors, from which we continue to benefit, and for which we are grateful.

The articles published in this Issue offer insight into a wide range of contemporary legal issues, including liability in respect of artificial intelligence and automated systems; the place of digital assets within the framework of property law; women's rights within Ghana's marital property regime; the proportionality test as applied in Indian constitutional law; the approach taken by Austrian and German courts to what might loosely be termed 'political' questions; and the protection of sound marks and phonograms in EU intellectual property law. Perhaps closer to home, this Issue also discusses the implications of the Retained EU Law (Revocation and Reform) Bill—colloquially known as the 'Brexit Freedoms Bill'—which is currently making its way through Parliament at the time of publication and has been the subject of considerable public debate. The breadth and depth of the articles published will hopefully make them of interest to British and international, student and professional audiences alike. The international and comparative nature of many of the articles in this Issue also serve as a reminder that there is much to be gained from an understanding of the laws of other jurisdictions.

We would like to express our gratitude to the Editorial Board for their work in reviewing and editing submissions, especially to our team of International Editors for providing comments and guidance in respect of submissions pertaining to jurisdictions other than England and Wales.

We look forward to presenting the Autumn Issue later this year.

Leo Pang and Sebastian Aguirre
March 2023

Table of Contents

Cambridge Law Review (2023) Vol 8, Issue 1, 1–24

The Guilty (Silicon) Mind: Blameworthiness and Liability in Human-Machine Teaming

BRENDAN WALKER-MUNRO* AND ZENA ASSAAD**

ABSTRACT

As science pushes the boundaries of the development of artificial intelligence (AI), the progress has caused scholars and policymakers alike to question the legality of utilising AI in various human endeavours. Debate has raged in international scholarship about the legitimacy of applying AI to weapon systems to form lethal autonomous weapon systems (LAWS). Yet the legality of applying or utilising AI is questionable even when AI is applied to a non-weaponised autonomous system: how does one hold a machine accountable for a crime? What about a tort? Can an artificial agent understand the moral and ethical content of its instructions? These are thorny questions, and in many cases, these questions have been answered in the negative, as artificial entities lack any contingent moral agency. What then occurs if the AI is not alone, but linked with or overseen by a human being, with their own moral and ethical understandings and obligations? Who is responsible for any malfeasance that may be committed? Does the human bear the legal risks of unethical or immoral decisions of an AI? These are some of the questions with which this manuscript seeks to engage.

Keywords: human-machine teaming, liability, criminal law, civil law, military

I. INTRODUCTION

Automation has been a key result of mankind's technological development over the last two centuries. Rather than a reliance on manual labour, we have developed mechanised tools which replace our efforts with streamlined and optimised acts.

* Senior Research Fellow, Law and Future of War Research Group, The University of Queensland; JD, PhD.

** Senior Research Fellow, School of Engineering, Australian National University; BAeroEng, PhD.

The research for this paper received funding from the Australian Government through Trusted Autonomous Systems, a Defence Cooperative Research Centre funded through the Next Generation Technologies Fund.

Even in the most sensitive and value-driven theatre of human endeavour—that of decision making—the march of progress has not slowed, such that we now have computer programs capable of making decisions on everything from restaurant orders and hotel bookings to delivery of healthcare and social welfare programs.[1]

Yet that automation is not without its controversy. A discussion has raged in the international community regarding the legitimacy of merging the 'hard' processing capabilities of a computer with the 'soft' processing abilities of a human.[2] Whilst the reality of such a concept might have been previously restricted to the pages of popular fiction,[3] this is no longer the case. Human-machine interfaces—where a system operates to modulate a human's sensory connection with a machine—are already being used in contemporary applications such as piloting drones and other autonomous and semi-autonomous platforms.[4] Scholars are now examining the next step of this inclusion of machines in the human realm of decision-making with an increased research interest in 'human-machine teaming' (HMT).[5] For the context of this paper, HMT is defined as a bi-directional combination of human and machine capabilities which work together with a dynamic directedness towards an aligned goal.[6]

Conceptually, this definition requires an HMT to include the processing capabilities of both a machine and human component. Each component must be able to send and receive messages from the other that enable them to actively (not passively) aim towards achieving the same goal. The concept of 'dynamic directedness' thus imports two requirements to HMT: one, an element of back-and-forth communication between machine and human component; and two, the need for that back-and-forth communication to be directed towards a similar (but not identical) objective purpose.

Yet despite the research interest in HMT, the literature still lacks a cohesive framework which adequately reflects the legal responsibility for an HMT. Imagine a car assembly, where human workers and robotic workers complete their tasks

[1] Igor Bikeev and others, 'Criminological Risks and Legal Aspects of Artificial Intelligence Implementation' (Proceedings of the International Conference on Artificial Intelligence, Information Processing and Cloud Computing, Sanya, December 2019) <https://www.researchgate.net/publication/337883901_Criminological_risks_and_legal_aspects_of_artificial_intelligence_implementation> accessed 22 March 2023.

[2] Linda Skitka, Kathleen L Mosier, and Mark Burdick, 'Does Automation Bias Decision-Making?' (1999) 51 International Journal of Human-Computer Studies 991; Ericka Rovira, Kathleen McGarry, and Raja Parasuraman, 'Effects of Imperfect Automation on Decision Making in a Simulated Command and Control Task' (2007) 49 Human Factors 76; Gustav Markkula and others, 'Models of Human Decision-Making as Tools for Estimating and Optimizing Impacts of Vehicle Automation' (2018) 2672(37) Transportation Research Record 153; Monika Zalnieriute, Lyria Bennett Moses, and George Williams, 'The Rule of Law and Automation of Government Decision-Making' (2019) 82 Modern Law Review 425.

[3] Alan Turing, 'Computing Machinery and Intelligence' (1950) 59 Mind 433.

[4] Jennifer Riley and others, 'Situation Awareness in Human-Robot Interaction: Challenges and User Interface Requirements' in Michael Barnes and Florian Jentsch (eds), *Human-Robot Interactions in Future Military Operations* (CRC Press 2017) 180.

[5] The terms 'human-machine team' and 'human-machine teaming' are functionally the same for present purposes and can be used interchangeably throughout this paper.

[6] Adapted from Memunat A Ibrahim, Zena Assaad, and Elizabeth Williams, 'Trust and Communication in Human-Machine Teaming' (2022) 10 Frontiers in Physics 1.

side-by-side, assembling the components of a vehicle as part of a smoothly operating team. Both humans and machines, however, are also given a particular values framework imposed by the factory owner: vehicles must be completed to a certain standard and within a certain time. What happens when the machines realise that their human counterparts are the ones that are slowing down the process, making mistakes, and costing time and resources? A human worker might seek to disobey the restrictions imposed on them by the factory owner, go on strike, or maybe just go at their own pace and risk dismissal. Robots, programmed by humans, might be programmed to behave by them. Alternatively, they may lack flexibility in the programming and kill their co-workers inadvertently in the pursuit of improvement.

Although this may sound like the plot to a Hollywood blockbuster, some semblance of these facts can be found in reality. Kenji Urada is widely recognised as the first human to die from an injury caused by a robot. In 1981, Urada was performing maintenance on an automated hydraulic arm which, despite written safety protocol, was still powered on. The system misinterpreted Urada's actions as an attempt to damage the arm, which reacted by knocking Urada into an adjacent machine. Urada was crushed and died instantly.[7] A similarly horrifying (though less serious) incident occurred in 2022 when a 7-year-old chess player had his finger broken by a robotic opponent.[8] In both cases, blame was laid squarely on Urada and the 7-year-old (that is, the human) for violating safety protocol, and otherwise no charges were laid.

Nowhere should this development be more concerning than in the field of military and armed forces given the rapid development of research into the 'deployment of AI-infused systems (e.g. drone swarming, command and control decision-making support systems and a broader range of autonomous weapon systems)'.[9] Whilst the idea of an HMT presents obvious benefits to military operations, the controversy arises by inflaming existing risks or generating new challenges of relevance to military commanders and systems designers. There should be a conceptual question about the attribution of responsibility for unlawful actions committed within military HMT operations: do the actions give rise to civil liability (where the remedy is usually compensation or some remedial order of the court) or criminal liability (where the remedy is usually imprisonment for natural persons, both as a form of punishment and to protect innocent members of society)?

[7] Yueh-Hsuan Weng, Chien-Hsun Chen, and Chuen-Tsai Sun, 'Toward the Human-Robot Co-Existence Society: On Safety Intelligence for Next Generation Robots' (2009) 1 International Journal of Social Robotics 267, 273.

[8] Jon Henley, 'Chess Robot Grabs and Breaks Finger of Seven-Year-Old Opponent' *The Guardian* (London, 24 July 2022) <www.theguardian.com/sport/2022/jul/24/chess-robot-grabs-and-breaks-finger-of-seven-year-old-opponent-moscow> accessed 16 September 2022.

[9] James Johnson, 'The AI-Cyber Nexus: Implications for Military Escalation, Deterrence and Strategic Stability' (2019) 4 Journal of Cyber Policy 442.

We, therefore, set out in this article to advance the proposition that, for military HMT operations, the specifics of the dynamic interactions between the human and machine elements will dictate how liability will be attributed. We intend to approach the problem in the following way. Section II will involve an exploration of the issues of defining HMT in the contemporary context. It will identify that the bi-directionality of communication between the human and machine elements serves to blur the perceptions and observations of both parts, and may have legal and regulatory ramifications. Section III will then introduce some key terms in the context of both civil and criminal law around the establishment of liability, with reference to the idea of blameworthiness. In Section IV, we explore some of the problems with applying these concepts of liability to an artificial agent (including a partial agent such as would be present in an HMT). Before concluding the paper, in Section V, we introduce and explain a possible mechanism of regulating an HMT which we argue will appropriately respect the concepts of blameworthiness and liability whilst retaining utility to incorporate artificial agents.

This article will also specifically focus on an HMT in a military context. There are three reasons for such a focus. The first is that an HMT is a significant component of the technological research for many western military forces including the US, the UK, the EU and Australia,[10] but also of west-adversarial nations such as China and Russia.[11] Secondly, like their comparative cousins in the form of autonomous weapon systems, the application of AI to military decision making in HMT is already recognised as a challenge to the rules-based order of international and comparative domestic law.[12] And thirdly, the military is often a testbed for emerging technologies, with the armed forces standing as the entity which most commonly responds to the legal and regulatory challenges that arise from their implementation.[13]

[10] UK Ministry of Defence, 'Joint Concept Note 1/18: Human Machine Teaming' (UK Ministry of Defence May 2018) <www.gov.uk/government/publications/human-machine-teaming-jcn-118> accessed 19 October 2022; Chad C Tossell and others, 'Appropriately Representing Military Tasks for Human-Machine Teaming Research' in Constantine Stephanidis, Jessie YC Chen and Gino Fragomeni (eds), *HCI International 2020 — Late Breaking Papers: Virtual and Augmented Reality* (Springer 2020); Alex Neads, David J Galbreath, and Theo Farell, 'From Tools to Teammates: Human Machine Teaming and the Future of Command and Control in the Australian Army' (Australian Army Occasional Paper No 7, 20 September 2021).

[11] US Department of Defense, 'Military and Security Developments Involving the People's Republic of China 2021' (US Department of Defense 2021) <https://media.defense.gov/2021/Nov/03/2002885874/-1/-1/0/2021-CMPR-FINAL.PDF> accessed 17 October 2022.

[12] Aiden Warren and Alek Hillas, 'Lethal Autonomous Weapons Systems: Adapting to the Future Unmanned Warfare and Unaccountable Robots' (2017) 12 Yale Journal of International Affairs 71; Aiden Warren and Alek Hillas, 'Friend or Frenemy? The Role of Trust in Human-Machine Teaming and Lethal Autonomous Weapons Systems' (2020) 31 Small Wars & Insurgencies 822.

[13] See for example how drone regulation has emerged in military contexts: Ferran Giones and Alexander Brem, 'From Toys to Tools: The Co-Evolution of Technological and Entrepreneurial Developments in the Drone Industry' (2017) 60 Business Horizons 875; Matthieu J Guitton, 'Fighting the Locusts: Implementing Military Countermeasures against Drones and Drone Swarms' (2021) 4 Scandinavian Journal of Military Studies 26.

II. Definitional Issues of Human-Machine Teaming

One of the most significant challenges facing the academic and industrial community is the lack of a shared definition of what exactly comprises an HMT. Definitions are vitally important for legal and regulatory purposes, not just as academic or theoretical constructs. The blurring of responsibility between the human and machine elements in an HMT and indeed the very concept of identifying where a human ends and a machine begins can present significant challenges to the legal and regulatory framework for future HMT operations. If a legal principle cannot apply to the emergence of HMT operations, or applies weakly or ambiguously, the danger of an unregulated system is apparent. Even absent the possibility that an HMT (especially military HMT) might be operating without a proper form of legal control or oversight, the absence of a proper regulatory system can diminish public trust in the operations of the armed forces. Worse, such systems could expose those same armed forces to liability themselves.[14]

One such example defines an HMT as 'a purposeful combination of human and cyber-physical elements that collaboratively pursue goals that are unachievable by either individually'.[15] Some broader literature on HMT have similarities in their proposed definitions, with many expressing notions of sharing authority to pursue common goals.[16] Such a definition clearly articulates the connection and bi-directionality between the human (natural) and the machine (artificial), yet articulates these by reference to a frame in which goals *cannot* be achieved by one or the other in isolation. Applying such a definition to the simple act of driving a vehicle highlights the definitional issues—clearly, both humans and machines can operate, steer, and control a vehicle without necessary recourse to the other.[17]

Another definition of HMT might be of more utility: 'the dynamic arrangement of humans and cyber-physical elements into a team structure that capitalizes on the respective strengths of each while circumventing their respective limitations in pursuit of shared goals'.[18] Is it the existence of collaboration then, of movement towards a shared goal, which hallmarks human-machine teaming? Yet again, the difficulty in the detail surfaces when applied to a contextual application. Imagine a drone equipped with missiles, deployed in a foreign state but monitored in its home state by a human operator. Both the drone and the operator have a shared goal—the identification, pursuit, and engagement of the State's legitimate military

[14] Consider for example the application of article 36 of Additional Protocol I of the Geneva Convention: Damian P Copeland, 'Legal Review of New Technology Weapons' in Hitoshi Nasu and Robert McLaughlin (eds), *New Technologies and the Law of Armed Conflict* (Springer 2014).

[15] Azad M Madni and Carla C Madni, 'Architectural Framework for Exploring Adaptive Human-Machine Teaming Options in Simulated Dynamic Environments' (2018) 6 Systems 44, 49.

[16] Joseph B Lyons and others, 'Human-Autonomy Teaming: Definitions, Debates, and Directions' (2021) 12 Frontiers in Psychology 1932.

[17] J Levinson and others, 'Towards Fully Autonomous Driving: Systems and Algorithms' (2011 IEEE Intelligent Vehicles Symposium IV, Baden-Baden, June 2011).

[18] Lyons and others (n 16).

targets—but the nature of the relationship is perhaps better characterised as supervision than 'circumventing their respective limitations'. The drone is obviously performing a function in replacement of the human operator and takes the risk in doing so, but the operator still is the one who carries the risk associated with commencing or prosecuting any attack.

How then do the various world militaries approach this definitional issue? The Australian Army broadly defines HMT as the 'incorporation of autonomous or robotic systems within military teams to achieve tactical outputs that neither machines nor people could deliver independently',[19] whilst the United Kingdom's joint concept note on HMT defines the 'effective integration of humans, artificial intelligence (AI) and robotics into warfighting systems'.[20] The US Department of Defense does not strictly define HMT, instead referring to it more obliquely via terminology buried in the program definitions. Take for example the Next-Generation Nonsurgical Neurotechnology (N3) project, which:

> …aims to develop high-performance, bi-directional brain-machine interfaces for able-bodied service members. Such interfaces would be enabling technology for diverse national security applications such as control of unmanned aerial vehicles and active cyber defense systems or teaming with computer systems to successfully multitask during complex military missions.[21]

What is common about these military definitions is the incorporation of, or integration between, human and machine components to achieve outcomes for the armed forces in combat and peacetime. These similarities in the military context also betray the same difficulties in the execution of HMT, which is to explain why military forces strive to achieve the ideal of HMT. The Australian army's definition contains perhaps the most succinct policy purpose of HMT, that is, to '…achieve tactical outputs that neither machines nor people could deliver independently',[22] a concept directly reflecting the ideal in the literature that HMT ought to circumvent the respective limitations of human and machine.[23]

A similar lack of consistency affects definitions in other research spheres. For example, the report published by the UN Institute for Disarmament Research (UNIDIR) does not explicitly define HMT. Instead, UNIDIR focuses on defining a spectrum of HMT operations from 'coactive design' to 'immersion' in which the

[19] Neads, Galbreath, and Farell (n 10).
[20] UK Ministry of Defence (n 10) 39.
[21] Gopal Sarma, 'Our Research: Next-Generation Nonsurgical Neurotechnology' (*DARPA*, 2022) <https://www.darpa.mil/program/next-generation-nonsurgical-neurotechnology> accessed 1 December 2022.
[22] Neads, Galbreath, and Farell (n 10).
[23] Madni and Madni (n 15); Lyons and others (n 16).

machine and human operate in 'virtual worlds that are simulated [and] dynamic'.[24] NATO approaches to HMT also focus not on the definition of the term, but on the supposed benefits to military decision-making, noting that teaming is the ultimate expression of collaboration, trustworthiness and adaptation between the human and machine components.[25]

A critical thread, however, can be observed across both the military and non-military works seeking to define HMT. Contemporary military capabilities already involve collaboration between humans and machines—whether the machine is a sensor, interface, weapon, or system capable of communicating in a shared language, to achieve or move towards some shared goal. The sharing of these capabilities between humans and machines is necessary for achieving outputs that neither entity could complete independently. The critical thread observed, however, in these definitions (and the one that lies at the core of this article) is the bi-directionality of that communication. A machine may communicate with its human capability by displaying sensor information, or the projected results of a weapon detonation,[26] whilst a human may provide commands to select, track or engage targets presented.[27]

Moreover, the bi-directionality of communication also presents a unique challenge to attributing responsibility to the agent in the team who made the particular decision. Autonomous weapons and military robotics have long been suggested to suffer from a 'responsibility gap'[28], that is, the idea that mankind cedes control to machines when they are invested with the capability to learn and

[24] Ioana Puscas, 'Human-Machine Interfaces in Autonomous Weapon Systems: Considerations for Human Control' (*UNIDIR*, 2022) 15–16 <https://www.unidir.org/publication/human-machine-interfaces-autonomous-weapon-systems> accessed 25 February 2023.

[25] Karel van den Bosch and Adelbert Bronkhorst, 'Human-AI Cooperation to Benefit Military Decision Making' (NATO Report STO-MP-IST-160 2018) 8 <https://karelvandenbosch.nl/documents/2018_Bosch_etal_NATO-IST160_Human-AI_Cooperation_in_Military_Decision_Making.pdf> accessed 25 February 2023.

[26] See for example the Athena AI which can differentiate between objects protected under international humanitarian law and legitimate targets: Jonathan Bradley, 'Athena AI Helps Soldiers on the Battlefield Identify Protected Targets' (*Create Digital*, 26 April 2021) <https://createdigital.org.au/athena-ai-helps-soldiers-identify-protected-targets/> accessed 13 July 2022.

[27] Vasja Badalič, 'Automating the Target Selection Process: Humans, Semiautonomous Weapons Systems, and the Assault on International Humanitarian Law' in Aleš Završnik and Vasja Badalič (eds), *Automating Crime Prevention, Surveillance, and Military Operations* (Springer 2021).

[28] Thomas Hellström, 'On the Moral Responsibility of Military Robots' (2013) 15 Ethics and Information Technology 99; Merel Noorman and Deborah G Johnson, 'Negotiating Autonomy and Responsibility in Military Robots' (2014) 16 Ethics and Information Technology 51; Lambèr Royakkers and Peter Olsthoorn, 'Lethal Military Robots: Who is Responsible When Things Go Wrong?' in Mehdi Khosrow-Pour (ed) *Unmanned Aerial Vehicles: Breakthroughs in Research and Practice* (IGI Global 2019); Bernd W Wirtz, Jan C Weyerer, and Carolin Geyer, 'Artificial Intelligence and the Public Sector—Applications and Challenges' (2019) 42 International Journal of Public Administration 596; Isaac Taylor, 'Who Is Responsible for Killer Robots? Autonomous Weapons, Group Agency, and the Military-Industrial Complex' (2021) 38 Journal of Applied Philosophy 320.

evolve, which Matthias postulated would lead to 'injustice of holding men responsible for actions of machines over which they could not have sufficient control'.[29] Ryan writes that 'military organizations must…examine whether it is desirable to have robots able to kill humans based on automated processes and without a human in the decision cycle'.[30] Nevertheless, the idea of a conjoined or collaborative HMT sidesteps Ryan's understanding of the issue: The existence of a human 'in the loop' of decision-making is no safeguard against failure.

Consider the following scenarios involving hypothetical military HMTs, but which have been based on existing automated or autonomous technologies:

1. The pilot of an attack aircraft, assisted by uncrewed sensor drones,[31] attacks a convoy based on the drones' assessment of those vehicles as being legitimate military targets. Following an investigation, it is revealed that the convoy contained fleeing refugees and the sensor data was incorrectly interpreted by the drones;
2. The captain of a Naval destroyer is linked to the automated defences of their ship. The radar detects an aircraft approaching and assesses its behaviour as benign; the captain, however, believes the aircraft is adopting an attack profile and opens fire. The aircraft was in fact an allied fighter in an adjacent battlegroup;[32] and
3. A platoon of soldiers is conducting a patrol in a foreign country, assisted by an armed robotic companion that is teamed with one of the platoon soldiers.[33] Unbeknownst to the platoon, the software underpinning the robot has been hacked by enemy forces and suddenly presents false threat warnings. The teamed soldier opens fire, killing one of his platoon members.

As shown above, each of these scenarios highlights a specific concern with the attribution of responsibility for a military HMT. In the first scenario, there was no malicious or adverse action by any person, merely the occurrence of what might

[29] Andreas Matthias, 'The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata' (2004) 6 Ethics and Information Technology 175, 183.

[30] Mick Ryan, 'Human-Machine Teaming for Future Ground Forces' (Center for Strategic and Budgetary Assessments 2018) 36 <https://csbaonline.org/uploads/documents/Human_Machine_Teaming_FinalFormat.pdf> accessed 25 February 2023.

[31] Based on the Loyal Wingman project developed by the Royal Australian Air Force: RAAF, 'Loyal Wingman' (*Royal Australian Airforce*, 2020) <https://www.airforce.gov.au/our-mission/loyal-wingman> accessed 9 July 2022.

[32] Adapted from the downing of a US intruder aircraft by a Japanese destroyer in 1996: Thomas Newdick, 'The Last Time a Japanese Warship Shot Down a U.S. Navy Plane was Actually not so Long Ago' (*The Drive*, 5 June 2021) <https://www.thedrive.com/the-war-zone/40937/the-last-time-a-japanese-warship-shot-down-a-u-s-navy-plane-was-actually-not-so-long-ago> accessed 14 January 2022.

[33] Based in part on the arming of a Boston Dynamics 'dog' robot: Joshua Rhett Miller, 'Robot Dog Equipped with Submachine Gun is 'Dystopian' Nightmare Fodder' *New York Post* (21 July 2022) <https://nypost.com/2022/07/21/robot-dog-with-submachine-gun-is-dystopian-nightmare-fodder/> accessed 17 August 2022.

be called 'human error'—yet it was an error that resulted in the preventable deaths of civilians.[34] In the second scenario, the naval captain imparted his human bias (contradicting the automated assessment of the aircraft's behaviour) into the decision-making cycle, arguably undermining the purpose of an HMT in the first place. In the last scenario, the addition of a cyber-physical element into human warfighting opens new avenues for misdirection and attack by enemy forces.

In all three scenarios, the issue of bi-directionality is front and centre in the difficulty of attributing responsibility. The pilot in the first scenario may well have been able to avert disaster had they not relied on the drones' sensor information and been able to visually observe the target, something required by pilots in previous conflicts.[35] In the second scenario, the communication with the human is what hampered the machine in (correctly) identifying that the aircraft was not a threat. Inversely, the third scenario demonstrates that the bi-directionality of communication in HMT introduces a vulnerability which can be exploited by adversarial forces. How might this bi-directionality then affect the legal treatment of HMT under civil or criminal law?

III. Liability in Civil and Criminal Law

At this point, it is apposite to examine the concept of liability in both civil and criminal law, so that the requisite characteristics of that concept can be identified which are vulnerable to displacement by HMT. This displacement is likely to occur because of the bi-directionality of communication between the human and machine elements of an HMT, and subsequent reliance on that communication as a basis for taking action: a decision made by a machine or human, where one influences the other, has the potential to affect resulting liability.

Traditional western legal systems attribute liability on a basis of 'the individual human person as the central unit of action and the appropriate object of blame'.[36] This idea of liability as blameworthiness, both factual and moral, informs how the civil and criminal look to achieve co-regulatory purposes by enforcing breaches of duties in ways that are generally complementary.[37] The criminal law attributes liability at a higher standard and burden of proof, acting in a more

[34] For example, consider the human error that led to the US strike on a Medecins Sans Frontiers hospital in Kunduz, Afghanistan in 2015: John F Campbell, 'Investigation Report of the Airstrike on the Medecins Sans Frontieres / Doctors Without Borders Trauma Center in Kunduz, Afghanistan on 3 October 2015' (United States Forces Command 21 November 2015) <http://fpp.cc/wp-content/uploads/01.-AR-15-6-Inv-Rpt-Doctors-Without-Borders-3-Oct-15_CLEAR.pdf> accessed 22 September 2022.

[35] As an example, the British Royal Air Force bombers colloquially known as 'Dambusters' in the Second World War were not permitted to drop their bombs until the dams were in sight: Paul Brickhill, *The Dam Busters* (Macmillan 2017).

[36] Neha Jain, 'Autonomous Weapon Systems: New Frameworks for Individual Responsibility' in Nehal Bhuta and others (eds), *Autonomous Weapons Systems: Law, Ethics, Policy* (Cambridge University Press 2016) 303.

[37] Kenneth W Simons, 'The Crime/Tort Distinction: Legal Doctrine and Normative Perspectives' (2008) 17 Widener Law Journal 719.

'agent-focused' manner than the civil law, where mere negligence or breach of duty will suffice.[38] Though criminal liability might involve a similar assessment of compliance or non-compliance as civil law,[39] the criminal law is also a tool of social control designed to both punish those who have committed the offences as well as to virtue signal potential future offenders that such behaviour is anathematic to good policy.[40] The stigma of criminal convictions and incarceration also achieves a broader social effect than the necessity of remediating or repairing harm in the context of civil litigation.[41]

In this way, crimes outlaw particular activity and make it impermissible under every circumstance, whilst civil law prevents the breaches of rights and provides reparation of breaches. In economic terms, 'criminal law exclusively imposes *sanctions*, while [civil] law prices an *activity*'.[42] The two are not mutually exclusive—sometimes civil law can be used to punish, whilst criminal law can be used to remediate.[43] Nevertheless, the importance of criminal and civil systems remaining complementary and co-regulatory, but existing as separate strands of law cannot be underestimated:

> …it is a mistake to compare crime and tort. If three persons are incited by a fourth to break into a house and cause damage each will be guilty of a crime and will receive separate punishment. The inciter will be guilty of the criminal offence of inciting others to commit crime. The other three will be guilty of the crime of breaking in. If the damage [is] caused…then in a civil action the three who caused the damage will be jointly and severally liable… The inciter will also be jointly and severally liable for the damage if he procures the commission of the tort and is a joint tortfeasor.[44]

Liability as blameworthiness is thus a common cornerstone to both civil and criminal law, even if they are crafted and applied in different contexts.[45] In civil law, blameworthiness is usually established by applying common law principles such as taking reasonable care not to harm one's 'neighbour' or person proximate to their conduct,[46] whereas for criminal law, it is the written Acts of some governing

[38] Peter Cane, 'Mens Rea in Tort Law' (2000) 20 Oxford Journal of Legal Studies 533, 555.
[39] Andrew von Hirsch and Martin Wasik, 'Civil Disqualifications Attending Conviction: A Suggested Conceptual Framework' (1997) 56 Cambridge Law Journal 599.
[40] Edoardo Greppi, 'The Evolution of Individual Criminal Responsibility Under International Law' (1999) 81 International Review of the Red Cross 531, 536–537. See also Rebecca Crootof, 'War Torts: Accountability for Autonomous Weapons' (2016) 164 University of Pennsylvania Law Review 1347.
[41] Crootof (n 40) 1361.
[42] Robert Cooter, 'Prices and Sanctions' (1984) 84 Columbia Law Review 1523. See also Cane (n 38).
[43] Cane (n 38).
[44] *CBS Songs Ltd v Amstrad Consumer Electronics plc* [1988] AC 1013 (HL).
[45] William J Stuntz, 'Substance, Process, and the Civil-Criminal Line' (1996) 7 Journal of Contemporary Legal Issues 1, 19.
[46] *Donoghue v Stevenson* [1932] AC 562 (HL).

body such as Parliament or Congress that set out the rules to be complied with.[47] In respect of making determinations of liability, the arbiter of law (the judge) and the arbiter of fact (often a judge but occasionally a jury) are called to offer an assessment of whether one party has broken a particular rule or breached a given duty.[48]

Given the further social significance of a criminal finding of guilt (potentially involving the loss of an individual's liability through a custodial sentence) versus the pecuniary imposition of damages through establishing negligence, the standard of proof for criminal liability is objectively higher than in civil law. This concept is expressed in most legal systems as beyond reasonable doubt as opposed to on the balance of probabilities,[49] and is expressed in somewhat equivocal terms in *Currie v Dempsey*:

> In my opinion [the legal burden of proof] lies on a plaintiff, if the fact alleged (whether affirmative or negative in form) is an essential element in his cause of action, eg if its existence is a condition precedent to his right to maintain the action. The onus is on the defendant, if the allegation is not a denial of an essential ingredient in the cause of action, but is one which, if established, will constitute a good defence, that is, an 'avoidance' of the claim which, prima facie, the plaintiff has.[50]

Moral and physical blameworthiness is also imported into other terms used in the determination of liability. Upon assessment of a particular factual situation, questions may be asked around the intent to engage in a particular act, which in turn invoke determinations of whether an action involves 'strict' liability or whether liability is contingent upon finding a person holding a particular state of mind—legally, the *mens rea* or 'guilty mind'.[51] It is only after exploring the complete factual situation that a person can be held responsible for some kind of illegal or wrongful act.[52]

This determination involves the importation of concepts of knowledge and intention to constitute moral blameworthiness, responsibility, and punishment.[53] Put differently, the concept of intent provides for the ascription of blameworthiness, a reflection of the aphorism that 'an agent is responsible for all and only his intentional actions'.[54] Collectively, lawyers commonly talk of intent as both a mental state of intending some action, and intentionality of the action as motivated by

[47] Douglas Husak, *Overcriminalization: The Limits of the Criminal Law* (Oxford University Press 2008) 9–10.
[48] Mike Redmayne, 'Standards of Proof in Civil Litigation' (1999) 62 Modern Law Review 167.
[49] In Australia, this is discussed in the seminal case of *Briginshaw v Briginshaw* (1938) 60 CLR 336.
[50] *Currie v Dempsey* (1967) 69 SR (NSW) 116, 539.
[51] Matthew R Ginther and others, 'The Language of Mens Rea' (2014) 67 Vanderbilt Law Review 1327.
[52] *Vines v Djordjevitch* (1955) 91 CLR 512, 519.
[53] Bertram F Malle and Sarah E Nelson, 'Judging Mens Rea: The Tension between Folk Concepts and Legal Concepts of Intentionality' (2003) 21 Behavioral Sciences and the Law 563, 564.
[54] John L Mackie, *Ethics: Inventing Right and Wrong* (Penguin UK 1990) 208.

that mental state.[55] Intentionality in criminal law has a defined and precise meaning and purpose, consisting of both the intention to engage in certain conduct and an intention to bring about a result because of that conduct (or knowledge that it will occur).[56] This is a deliberate choice: though 'strict' liability exists in crime where no proof of intention is needed, it is usually reserved for minor or regulatory offences where the removal of proving intent is not considered procedurally unfair to the accused.[57] Equally, punishing only those offences that a person actually plans and then carries out severely constrains the legal system in regulating unlawful conduct.[58]

Hence, although intentionality and intention may appear similar in both civil and criminal law, they are treated differently and can achieve different outcomes. Good motives cannot rescue or defend wrongful conduct, either in tort or crime. In *Caldwell*[59] an individual erected a wharf on public property and was charged with public nuisance. His defence—that the wharf was at the request of, and benefitted, the local community—was dismissed by the court because he had infringed a common right. On the other hand, a malign motive will taint any form of conduct, even if the conduct itself is morally acceptable. For example, a contract is a lawful arrangement between two parties and may be undertaken by any persons in society at large to regulate their dealings. However, a contract that is objectionable on public policy or legal grounds—such as a contract to commit murder—is void and unenforceable.[60]

Thus, criminal law departs from civil law because the bare formulation of mental state and conduct grounds liability, and there is no need to prove a particular effect or outcome. This explains the criminalisation of conduct even where both parties may consent (such as drug dealing or prostitution[61]), where the offence never actually took place (such as attempting to commit a crime[62]) or where the offence was committed by someone else (inchoate crimes such as aiding or abetting, which are treated differently to contributory negligence[63]). Further, it is almost always the State—and not the infringed party—who brings proceedings for the commission of crimes.[64] Conduct might also be criminalised without reference

[55] Malle and Nelson (n 53).
[56] Issues of automatism and involuntariness are beyond the scope of this paper.
[57] David Prendergast, 'The Constitutionality of Strict Liability in Criminal Law' (2011) 33 Dublin University Law Journal 285. Cf Federico Picinali, 'The Denial of Procedural Safeguards in Trials for Regulatory Offences: A Justification' (2017) 11 Criminal Law and Philosophy 681.
[58] Cane (n 38) 553.
[59] *Respublica v Caldwell* 1 US (1 Dall) 150 (Pa Ct of Oyer & Terminer 1785).
[60] *Commonwealth Bank of Australia Ltd v Amadio* (1983) 151 CLR 447.
[61] Barbara Sullivan, 'Rape, Prostitution and Consent' (2007) 40 Australian & New Zealand Journal of Criminology 127.
[62] Ian D Leader-Elliott, 'Framing Preparatory Inchoate Offences in the Criminal Code: The Identity Crime Debacle' (2011) 35 Criminal Law Journal 80.
[63] Joachim Dietrich, 'The Liability of Accessories under Statute, in Equity, and in Criminal Law: Some Common Problems and (Perhaps) Common Solutions' (2010) 34 Melbourne University Law Review 106.
[64] Ric Simmons, 'Private Criminal Justice' (2007) 42 Wake Forest Law Review 911.

to culpability if there was a serious social cost. Referencing Blackstone's *Commentaries*, Binder observed that the formulation of early Crown offences such as treason, carnal knowledge of the queen, piracy, serving a foreign monarch, or harbouring a Catholic priest were punishable without any proof of intent.[65]

Conversely, the purpose of proving intention in civil law (especially torts)—as opposed to in criminal law, where intention may be a fundamental proof of the charge—may be unnecessary. Torts are almost always actioned by the aggrieved parties, and not the State, to receive remedies that place the aggrieved parties as near to their original position before the infringement.[66] Because the focus of tort liability is generally on the existence of a duty of care, a breach of that duty, and in most cases, the suffering of harm, one cannot attempt a tort, plan one, or conspire to cause one.[67] Intention is usually relevant to penalty, not liability; again, this is a deliberate choice. For the victim whose rights have been infringed, they might not necessarily care if an infringement was actuated by malice, recklessness, or negligence. A search for intentionality may well be meaningless to compensate the victim for the harm suffered.

That is not to say that intention in civil law is a useless concept. Exemplary damages may be issued by the court in cases where the conduct was deliberately engaged in and 'of a sufficiently reprehensible kind'.[68] In this way, torts can punish intentional conduct in circumstances where an 'assertion of one's autonomy… produces harmful consequences [it] may justify more onerous liability than negligence'.[69] Intention is also more relevant where torts regulate activity with high social value high risk, such as transporting dangerous goods or manufacturing poisonous chemicals. In these contexts, it is apparent that the differences between negligence and malice are far more relevant to tortious conduct. In the words of Cane, 'when a harm-causing activity has high social value, a requirement of intention for tort liability helps to protect society's interest in the continuance of that activity'.[70]

IV. Problems Applying Liablity to HMT

The appropriate and proportionate imposition of liability in the context of artificial agents is not a novel problem to confront the law. The European Union (EU)

[65] Guyora Binder, 'The Rhetoric of Motive and Intent' (2002) 6 Buffalo Criminal Law Review 16.
[66] Scott Hershovitz, 'The Search for a Grand Unified Theory of Tort Law' (2017) 130 Harvard Law Review 942. Cf Seth Davis and Christopher A Whytock, 'State Remedies for Human Rights' (2018) 98 Boston University Law Review 397.
[67] Though a party may face contributory negligence for playing some part in its commission: Paul S Davies, 'Accessory Liability for Assisting Torts' (2011) 70 Cambridge Law Journal 353; Paul S Davies and Philip Sales, 'Intentional Harm, Accessories and Conspiracies' (2018) 134 Law Quarterly Review 69.
[68] *Lamb v Cotogno* (1987) 164 CLR 1.
[69] Cane (n 38) 553.
[70] ibid.

for example has proposed an Artificial Intelligence Act, the first one like it anywhere in the world.[71] Such regulation would provide a broader contextual and developmental framework for the design and implementation of AI systems, with the European Commission already having adopted a directive regarding the liability of AI systems.[72] Whilst the EU Directive only deals with civil and not criminal liability, it does emplace liability markers on designers, manufacturers, testers and end-users of AI systems where those systems do not comply with the principles in the proposed AI Act.

This places the EU at the forefront of regulating AI, but other countries are aware of the regulatory impacts of AI. The US has passed several legislative instruments which impose obligations upon government agencies to develop rules and guidelines for the design and testing of AI systems, especially for making decisions in the government.[73] By contrast, the UK intends to make no global rules governing AI, but to leave regulation down to the existing regulators such as the Competition and Markets Authority, the Information Commissioner's Office, and the Financial Conduct Authority.[74]

One of the key threads linking all these proposals is that military use of AI—whether meeting the terms of our proposed definition of HMT or otherwise—is excluded, either explicitly or by implication. Further, all of these legislative proposals do little to oust existing rules of liability, where States are largely able to invoke the protection of the 'act of State' doctrine to prevent courts from imposing liability on defence decisions.[75] The EU AI Act explicitly carves out military use whilst the EU Directive provides that national courts must limit disclosure and preserve secrecy in cases involving potentially confidential evidence (which one presumes would cover the technical specifications of military technology). In the US, it is the Office of the Under Secretary of State for Arms Control and International Security that provides for military AI regulation, yet the agency

[71] Noting that the proposed regulation 'shall not apply to AI systems developed or used exclusively for military purposes': Commission, 'Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts' COM (2021) 206 final (EU AI Act).

[72] Commission, 'Proposal for a Directive of the European Parliament and of the Council on Adapting Non-Contractual Civil Liability Rules to Artificial Intelligence' COM (2022) 496 final (EU Directive).

[73] Namely, the National Artificial Intelligence Initiative Act of 2020 (Division E, § 5001) and the AI in Government Act of 2020 (Division U, Title I). See also Executive Order 13960, 'Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government' (Federal Register 12 August 2020) <https://www.federalregister.gov/do-cuments/2020/12/08/2020-27065/promoting-the-use-of-trustworthy-artificial-intelligence-in-the-federal-government> accessed 25 February 2023.

[74] Department for Digital Culture, Media, and Sport, 'Establishing a Pro-Innovation Approach to Regulating AI: An Overview of the UK's Emerging Approach' (Department for Digital Culture, Media, and Sport 18 July 2022) <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1092630/_CP_728__-_Establishing_a_pro-innovation_approach_to_regulating_AI.pdf> accessed 20 January 2023.

[75] Brendan Walker-Munro, 'Exploring Manufacturer Strict Liability as Regulation for Autonomous Military Systems' (2022) 27 Torts Law Journal 182. Compare the scope of the defence in the US courts in *Boyle v United Technologies Corp* 487 US 500 (1988) to the English courts in *Smith v Ministry of Defence* [2013] UKSC 41, [2014] AC 52.

is completely omitted from US AI legislation. The UK carves out military applications under its proposed policy by explicitly mentioning the Ministry of Defence as being a 'domain that [has] existing and distinct approaches to AI regulation' and therefore permitted to write its own ruleset. At the global level, the Group of Governmental Experts under the Convention on Certain Conventional Weapons have argued for seven years about the definitional scope of AI in the military.[76] It is appropriate then that we consider some of the challenges to these existing models of legal regulation that are being applied in the sphere of AI and how they may not sit easily in the context of a military HMT.

A. Theoretical challenges to HMT liability

It is this focus on blameworthiness that will likely be disrupted by the appearance or adoption of HMT and its bi-directional communication between man and machine. In a legal system where the focus is on the punishment of unlawful conduct or the remediation of breaches of rights, any circumstance influencing the blameworthiness of an agent will have serious ramifications for attribution of liability:

> [A]n agent can only be held responsible if they know the particular facts that surround their action, they are able to freely form a decision to act, and are able to select one of the suitable available alternative actions based on the facts of the given situation.[77]

Breaking apart this statement, we can consider three consecutive notions of attribution of liability and blameworthiness that are worth further exploration in the context of HMTs: *knowledge* of the facts, the existence of *suitable alternatives*, and the *freedom* to decide on one of them:[78]

1. Knowledge: Consider for a moment an HMT where the machine component of the team merely provides information or feedback to the human component, but the machine's programming suffers a catastrophic error and the feedback the human receives is completely nonsensical. A decision to engage in some form of conduct based on that erroneous information might be actionable if the conduct is subsequently proven to be unlawful;
2. Suitable alternatives: Such an assessment is in part subjective and in part objective, considering what the individual thought suitable as

[76] Klaudia Klonowska, 'Article 36: Review of AI Decision-Support Systems and Other Emerging Technologies of Warfare' (2022) 23 Yearbook of International Humanitarian Law 123.
[77] Matthias (n 29) 175.
[78] Giovanni Sartor and Andrea Omicini, 'The Autonomy of Technological Systems and Responsibilities for their Use' in Nehal Bhuta and others (eds), *Autonomous Weapons Systems: Law, Ethics, Policy* (Cambridge University Press 2016) 62.

well as the broader social context in which the act occurred.[79] Of course a lack of blameworthiness because of no suitable alternatives does not always exculpate the actor, as acts which are 'the lesser of two evils' can still be an infringement on another's rights;[80] and

3. Freedom of choice: Whether the human component of the HMT experienced a removal of freedom of choice will depend on the circumstances of the conduct and the context of the teaming operation. In circumstances of extreme emergency, or where the human and machine components are inextricably combined, there may be no way to divorce the human in any way that would render a valid freedom of choice. In others, the factual circumstances are highly relevant to blameworthiness: a military HMT operation in a combat zone may result in far less freedom of choice than an administrative HMT operation within an office.

By examining and weighing all three concepts we suggest it is possible to assess the degree to which an HMT might be liable for acts undertaken, and appropriately adjust for the artificial component's effect on human decision-making. The actions within an HMT, assessed partly on actions by a machine and partly on actions by a human, will become enmeshed to varying extents and may lead to overlapping spheres of liability for blameworthiness. Given the nature of HMTs, it is perhaps easiest to 'conceive of their actions as creating a web of overlapping chains of responsibility, both criminal and civil in nature'.[81] As a concept, this idea already appears in the literature in the context of attributing liability to autonomous systems more generally:

> The *mens rea* of the direct perpetrator therefore must be judged in terms of the secondary party's mental state, and will require intent or knowledge. This can also apply to AWS, as their code gives them the ability to perform some decision-making capabilities, and therefore be able to comprehend certain elements of their actions. However, ultimately, their actions are limited by a human agent, who sets parameters for how they are able to act. Therefore, responsibility can be shared by both the AWS and another human counterpart who is involved in its behaviours and actions.[82]

[79] Randolphe Clarke, 'Moral Responsibility, Guilt, and Retributivism' (2016) 20 The Journal of Ethics 121, 124.
[80] James Goudkamp, 'The Spurious Relationship Between Moral Blameworthiness and Liability for Negligence' (2004) 28 Melbourne University Law Review 343.
[81] Jain (n 36) 304.
[82] ibid 310.

B. Pragmatic challenges to HMT liability

There are also some unresolved difficulties in the application of liability and blameworthiness to HMT more generally. The first is identifying which actor within an HMT, whether the machine or human actor, is the one 'making' a decision when the tasks being completed are not repetitive or deterministic.[83] Consider the theoretical effects explored above: what if a human is presented with a tactical scenario in which there are no alternative options which the human considers acceptable? If the human takes what is the only 'reasonable' option, are they really making a decision? The decision has already been made by the machine—perhaps inadvertently—by presenting the information in a way that only one option was possible.

The second challenge, particularly in the military and armed forces context, is the effect of HMTs on the inquisitorial process (such as criminal investigation or civil discovery). These processes often involve determining both a factual substrate of the conduct, but also an assessment of liability. Unfortunately, HMT presents two distinct barriers to these processes. Firstly, much of the technology, automation, or software underpinning HMTs is likely to be protected by trade secrets or military secrecy;[84] and secondly, the opacity of AI-automation programs in HMT means that even where such the code of such programs can be exposed, the apparent nature of decision-making by that code is not readily discernible in a manner understandable by jurors or judges.[85]

The third is the differing legal treatment of various mental defences within and across jurisdictions. It is not within the scope of this article to consider the various natures of impairment, automatism or insanity defences (however they might be labelled); instead, it is to note that the varying degrees, scope, and application of these defences will lead to entirely varied treatments of HMTs in circumstances where judges are called to assess the 'voluntariness' of actions to assign blameworthiness.[86] This is especially the case where many of the mental defences often involve some level of 'impairment' to functioning. Does the human in an HMT really become 'impaired' because of the inclusion of a machine component?[87] Again, the machine may have removed the scope of involuntariness merely by presenting the information to the human in a particular format or fashion.

[83] Shagun Jhaver and others, 'Human-Machine Collaboration for Content Regulation: The Case of Reddit Automoderator' (2019) 26(5) ACM Transactions on Computer-Human Interaction 1, 8; Vincent Boulanin and others, *Limits on Autonomy in Weapon Systems: Identifying Practical Elements of Human Control* (Stockholm International Peace Research Institute 2020).

[84] Gabriele Spina Ali and Ronald Yu, 'Artificial Intelligence Between Transparency and Secrecy: From the EC Whitepaper to the AIA and Beyond' (2021) 12(3) European Journal of Law and Technology 1, 6–8.

[85] Ashley Deeks, 'The Judicial Demand for Explainable Artificial Intelligence' (2019) 119 Columbia Law Review 1829.

[86] Peter Cane, 'Fleeting Mental States' (2002) 59 Cambridge Law Journal 273.

[87] These defences are generally exculpatory doctrines to apply in only the most extreme cases: Steven Yannoulidis, 'Excusing Fleeting Mental States: Provocation, Involuntariness and Normative Practice' (2005) 12 Psychiatry, Psychology and Law 23, 27.

The last challenge for regulating HMTs in a pragmatic sense is determining a remedy that adequately reflects the blameworthiness of the conduct. Most western legal systems have evolved from the perspective that irrespective of the legal entity a claim is brought against, there is nevertheless a 'human who decides whether or not to comply'.[88] For example, civil and criminal actions are often brought against companies as legal entities, but where the actions of those companies are often a sheeted home to individuals within them.[89] Where a machine can be attributed with blameworthiness, there comes the question of how to achieve a penalty or restitution in a manner that is relevant to the machine. Alternately, there is a question of how to apply a remedy to a human who may have had no conscious control of or over the actions they are now alleged to have engaged in.[90]

In summary, these various principles of theoretical and pragmatic challenges in the context of military HMTs can be accounted for. Whatever the intended scheme of regulation is proposed, we consider that it must be capable of addressing the difficulties of applying blameworthiness in the context of HMT operations generally, but also the military and armed forces context more specifically. We consider the best and most efficient approach to involve modifying an existing regulatory scheme to apply to the future use of military HMT operations.

V. A Proposed Framework for Liability in HMT

In this final Section, we propose the leveraging of a concept that has already been explored in the literature—chains of responsibility or 'COR'—as a mechanism for attributing liability in HMT. Originating in the logistics and supply chain industry, COR applies a proactive model of compliance to prevent road and freight accidents. COR legislation for heavy vehicles is already a feature of the legal landscape in Australia.[91]

It is with this framework in mind that we present a COR model for the HMT context in Table V.1. Along the vertical axis, Table V.1 charts the lifecycle of an HMT from conception and design, through manufacture and testing, to procurement and deployment (both domestic and foreign). At each stage of that lifecycle, those involved with HMT will carry responsibilities explored horizontally. These responsibilities are non-exhaustive and intended to provide a high-level example of the types of activities at each stage which are relevant in determining potential legal culpability from the use of HMT. Each of them has been derived

[88] Lawrence Lessig, 'The Zones of Cyberspace' (1996) 48 Stanford Law Review 1403, 1408.

[89] Michael Nietsch, 'Corporate Illegal Conduct and Directors' Liability: An Approach to Personal Accountability for Violations of Corporate Legal Compliance' (2018) 18 Journal of Corporate Law Studies 151.

[90] David Watson, 'The Rhetoric and Reality of Anthropomorphism in Artificial Intelligence' (2019) 29 Minds and Machines 417.

[91] Heavy Vehicle National Law (Queensland); as applied by the Heavy Vehicle National Law Act 2012 (Qld), s 4; Heavy Vehicle National Law Act 2013 (ACT), s 7; Heavy Vehicle (Adoption of National Law) 2013 (NSW), s 4; Heavy Vehicle National Law (South Australia) Act 2013 (SA), s 4; Heavy Vehicle National Law (Tasmania) Act 2013 (TAS), s 4; Heavy Vehicle National Law Application Act 2013 (Vic), s 4.

from the theoretical and pragmatic challenges to HMT liability in Section II and is intended to be read from the perspective of 'reasonable foreseeability' For example, legal and ethical advice in the conduct of military operations is best sought well before the first shot is fired, when advising how a conflict may be fought.[92] Thus, legal and ethical advice should be incorporated into the very design of the HMT.[93] To discharge their responsibilities at the manufacturing stage, those producing HMT interfaces and software should have rigorous testing regimes in place which are capable of detecting flaws and errors to a low tolerance (noting that the systems will ultimately be used in a warfighting capability[94]). Those involved in the manufacture of subcomponents will also need to meet these rigorous standards and be informed by the principal manufacturer of the potential risks.[95]

TABLE V.1

Chain of Responsibility for HMT			
Design Phase	Legal and ethical advice	Are system functions being automated appropriately?	Comply with international and domestic law
Manufacturing	Advise subcontractors of potential liability	Testing at or above industry standard	Fully investigate adverse incidents
Testing	Frequent and robust testing	Address any potential safety concerns, that is, with mandatory warnings	Safety must be 'such as persons generally are entitled to expect'
Contract Negotiation	Include, as contractual terms, the intended scope and operating environment	Disclose all possible safety issues	Include maintenance and upkeep cycles in contract
In-Service	Further testing compatible with intended operational environment	Be aware 'handover' to military authorities does not end liability	Operators must not operate systems until deemed competent
Domestic Deployment	Platform must operate consistent with domestic and international laws	Human decision-maker must be capable of intervening at any time	Systems must be operated in accordance with the manufacturer's instructions and training at all times
International Deployment	Be aware that international and domestic law will apply	Human decision-maker must be capable of intervening at any time	Consider IT security in overseas environments

The concept of COR outlined in Table V.1 is relatively simplistic: it imposes a primary, direct, and non-delegable duty on every person interacting with an

[92] Marcus Schulzke, 'Autonomous Weapons and Distributed Responsibility' (2013) 26 Philosophy & Technology 203, 209.
[93] Heather M Roff, 'The Strategic Robot Problem: Lethal Autonomous Weapons in War' (2014) 13 Journal of Military Ethics 211.
[94] Jai Galliott, 'The Soldier's Tolerance for Autonomous Systems' (2018) 9 Paladyn, Journal of Behavioral Robotics 124.
[95] Walker-Munro (n 75).

HMT to ensure the safety of their individual activities so far as is reasonably practicable.[96] At each level, from design, through manufacture and testing, to 'handover' to military authorities and eventual deployment in military operations, an HMT must be rigorously tested in all intended operational environments. Legal and ethical advice should be sought and incorporated into the design, manufacture, and testing stages. Such testing must be performed by both the manufacturers and military authorities, and testing performed at any specific stage should not be regarded as being conclusive. A 'cut-off' or similar system should always be included in any HMT that permits a human operator (or other person acting remotely) to deactivate the machine component in the event of a failure or incident. Any safety defects, issues, and injuries must be rigorously investigated and either remediated or repaired, or a mandatory warning provided in relation to conduct likely to cause that issue again. In both training and operational use, military commanders bear an additional non-delegable duty to ensure their staff are trained on HMTs and deemed competent in their use. In the absence of clear legal guidance to the contrary, principles of both domestic and international law should be deemed to always apply to the use of HMTs in operational military environments.

In the event of an accident or incident, an investigation is conducted that examines the entire logistic chain to determine where the duty was breached, and by which agent. Breaches of that duty of care may result in the commencement either of civil action (involving pecuniary penalties) or criminal offences (involving potential for penal sentences in severe cases). There exists a legitimate question as to how COR might address any of the theoretical or pragmatic challenges that HMT poses to existing civil and criminal liability approaches. It is therefore the focus of this Part of the article to demonstrate how COR could apply in the context of military HMTs.

A. Applying COR to theoretical challenges to HMT liability

It should be recalled that we examined three consecutive notions of attribution of liability and blameworthiness applying to HMTs: *knowledge* of the facts, the existence of *suitable alternatives*, and the *freedom* to decide on one of them.[97] How should COR apply to these three notions of liability?

Firstly, COR examines the nature of actions taken and decisions made up to and inclusive of the decision to engage in the impugned conduct. In such an *ex*

[96] Amanda Beesley, 'Improving Safety and Compliance, and Simplifying Enforcement – Recent Reforms to Australia's Heavy Vehicle Chain of Responsibility Laws' (International Symposium on Heavy Vehicle Transport Technology, Rotorua, 5 October 2016); Geoff Farnsworth and Jarrad McCarthy, 'Heavy Vehicle National Law Reform: New Approach to Chain of Responsibility Liability' (2016) 68 Governance Directions 41; Wonmongo Lacina Soro, 'Towards an Understanding of Financial Influences on Heavy Vehicle Safety Outcomes' (PhD thesis, Queensland University of Technology 2020).

[97] Sartor and Omicini (n 78).

ante examination, it is not just the blameworthiness of the ultimate decision that is determinative, but of each of the 'steps' that led up to its final execution. Consider the earlier example of a military HMT where the machine programming fails and the human is presented with nonsensical information. In this case, the application of COR's 'reasonably practicable' assessment of safety might determine that the nature of the manufacturer's pre-deployment testing was insufficient and that this was the blameworthy failure. Alternately, it might be a repairer who inserted a faulty component who bears the blame for the incident at hand. This concept of extended liability is certainly not unknown to the literature and reinforces the idea that delictual responsibility is not a pie—'[a]ll involved can theoretically take all the responsibility for the harm caused, and if unjustified, punished'.[98]

Secondly, whilst COR applies equally across the nature of military and non-military actors, it has the flexibility to consider the unique challenges of military service. The application of COR is based on precautions that are 'reasonably practicable' by reference to the controls available, the suitability of those controls, and the cost of controls proportional to the risk posed. Consistent with other approaches to applying liability to emerging military technologies, it is easier to control risks from a manufacturer's office or a designer's factory, places which are far removed from operations against the enemy in a foreign war zone.[99]

Thirdly, COR sidesteps many of the theoretical issues to liability attribution that might occur in the context of military HMT. Again, the curial search in COR is for 'reasonable practicability', not necessarily strict blameworthiness. In circumstances where a military force has not properly trained a human for HMT operations but could have easily done so with the resources and time it had available, underlying fault questions do not arise. The military force bears responsibility under COR and may be prosecuted or litigated accordingly. Military forces are already assessed for this level of compliance under most western work health and safety systems, suggesting that the level of adaptation required to adopt COR is unlikely to be onerous or disruptive to military operations.[100]

Fourthly, by adopting a less prescriptive system for attribution of blameworthiness, there is potential to avoid the injustice of liability being applied to persons not having sufficient control of machines or in circumstances where the machine has malfunctioned.[101] At the same time, COR renders irrelevant the need for militaries to scrutinise the role of humans in a decision cycle.[102] The focus of inquiry in situations of failure is on the reasonableness of safeguards enacted to protect against harms, not the actions of the individual HMT.

[98] Ross W Bellaby, 'Can AI Weapons Make Ethical Decisions?' (2021) 40 Criminal Justice Ethics 86, 96.
[99] *Smith* (n 75); Walker-Munro (n 75).
[100] Nick Turner and Sarah J Tennant, '"As Far as is Reasonably Practicable": Socially Constructing Risk, Safety, and Accidents in Military Operations' (2010) 91 Journal of Business Ethics 21; John A Casciotti, 'Fundamentals of Military Health Law: Governance at the Crossroads of Health Care and Military Functions' (2016) 75 Air Force Law Review 201.
[101] Matthias (n 29).
[102] Ryan (n 30).

Finally, COR is equally adaptable to the vast array of field environments in which modern militaries are prepared to operate. The reasonableness of safety precautions to avoid specific harms recognises that 'a safety measure that would be enough in one situation might be completely inadequate in another and excessive in a third set of circumstances'.[103] Therefore, what may be deemed acceptable to limit HMT risk in a training setting might be inadequate for foreign operations but excessive in joint or allied exercises. Acceptability is dependent entirely on the circumstances of the HMT deployment and the likelihood and magnitude of the risk being guarded against.

B. Applying COR to pragmatic challenges to HMT liability

In the same vein, COR has the potential to drastically limit or eliminate the pragmatic risks to the attribution of HMT liability. The application of COR to HMT operations, especially military operations, recognises the unique factual circumstances in each deployment of HMT and seeks to impose a sliding duty of reasonableness to determine whether liability should apply and to what degree.

Firstly, the idea of needing to determine which aspect of an HMT—human or machine—'made' a decision for any liable conduct is irrelevant. The focus of COR is not strictly limited to the liable conduct in question, but on all the antecedent decisions and circumstances along the chain leading to that conduct. Where a programming error which presents inaccurate or misleading data in an HMT is the causative agent, liability is still attributable to the human for not verifying the information using another technique or system. The fighter pilot who bombs a target without visually verifying and satisfying themselves of a target's validity (and instead relying on the automated or autonomous system) stands to carry some of the punishment or rectification for that fault.

Secondly, COR considers but does not rely on the mental element of each of the individual actors along the chain. Each individual actor has the same duty ('to limit risk to the extent reasonably practicable') but different capacities and methods of discharging that duty, in the same way as modern work health and safety laws operate in the military context.[104] Defences of automatism, involuntariness, or diminished capacity are relevant only to the extent that the actor can discharge the duty, not the existence of the duty. It further recognises that the liability of one individual of the chain may be contingent, or rely upon, the liability of others. The failure of a soldier relying on faulty data is influenced by and partly reliant upon the failure of a manufacturer to properly test the machine components.

[103] National Heavy Vehicle Regulator, 'Primary Duty Definitions' (*National Heavy Vehicle Regulator*, 2022) <https://www.nhvr.gov.au/safety-accreditation-compliance/chain-of-responsibility/the-primary-duty/primary-duty-definitions> accessed 20 January 2023.
[104] Turner and Tennant (n 100).

Thirdly, COR has the potential to move with new developments in technology, including the ability for machines to achieve 'artificial general intelligence'.[105] At such a point, machine components in HMTs may well be treated as moral actors with their own level of agency, at which point they become another link in the COR and their potential for blameworthiness becomes examinable. If a machine can achieve general intelligence in a manner that can be attributed to moral agency, there is no reason why its actions could not be examined through the lens of reasonable practicability for preventing harm.

The fourth benefit to the application of COR in military applications of human-machine teaming is the ability to attribute liability to specific human actors interacting with the system over time. As we stated above, liability is attributed to the human based on the scope and scale of their interaction with the AI system, not the level of their involvement in the blameworthy decision. In that way, COR applies a remedy for blameworthiness to a human actor on the basis of what reasonable steps could or should have been taken. Equally, COR also recognises the taking of reasonable steps against the realisation of harm as a partial or full defence to that liability, incentivising actors to behave in protective and alleviative ways.

Of course, COR has limited ability to counter the challenges of trade secrets, secrecy, or explainable AI. Indeed, these potentially introduce a challenge in the form of the broadly conceptualised 'state of the art' defence. This defence obviates responsibility in COR and similar regimes for defects that could not be detected by reasonably practicable testing available at the time of manufacturing or programming.[106] Many western legal systems have, however, grappled with like concepts for decades. In most cases, they involve the interpretation and application of rules of procedure which are dealt with at the level of individual courts or tribunals (including their military equivalents). There are already calls for action across multiple domains in respect of vesting arbiters of fact with appropriate powers of inquiry, coupled with broader education of the legal fraternity in concepts like AI.[107]

VI. Conclusion

There can be little doubt that although the integration of human and machine elements offers significant benefits to the armed forces, insufficient consideration has been given to how to regulate these integrations. Given the significance of decisions made in the context of military operations, which might involve the deaths of hundreds or thousands of people, we cannot leave the regulation of such events to mere chance and ambiguity. Nor does there appear to be much benefit in

[105] Ben Goertzel, 'Artificial General Intelligence: Concept, State of the Art, and Future Prospects' (2014) 5(1) Journal of Artificial General Intelligence 1; Pei Wang, 'On Defining Artificial Intelligence' (2019) 10(2) Journal of Artificial General Intelligence 1.

[106] Mabel Tsui, 'The State of the Art Defence: Defining the Australian Experience in the Context of Pharmaceuticals' (2013) 13(1) QUT Law Review 132.

[107] Deeks (n 85).

merely outlawing the pursuit of HMT applications, driving the research underground, and delegitimising a purposeful line of human research.

Instead, what is required is a nuanced and purposeful regulatory regime which considers the reasoning for attribution of responsibility, whilst also providing appropriate mechanisms for restitution and punishment. This is much for the benefit of our armed forces as for the protection of the rules-based global order: military officers and personnel need to know the legal limits of their conduct, what can be done in war and peacetime, and what consequences might attach when they step outside those boundaries.

There is still more to be done. The exact parameters of technologies designed to constitute HMT and how they are defined in law will need a more comprehensive examination than was possible in this article. The definitions will need to be expansive enough to capture those technologies at the forefront of military and civilian research, but also those yet to be contemplated. Alternately, new legal definitions for those technologies will need to be included in their own regulatory regime to eliminate grey areas and ambiguity. Just like our treatment of AI, we need to ensure that the definition is clear, unambiguous, and is not leading to inaccurate or oversimplified definitions of the technology.[108]

The work on regulating HMT also will not end with the possible introduction of a COR regime. There will no doubt be developments in warfighting technology which escape even the most carefully drafted working definitions of HMTs. Systemic difficulties which cannot be resolved at the procedural level of courts and tribunals will inevitably arise. Future avenues of research might look at how COR regimes could be tailored to specific military operations, or how military COR might be adapted to civilian environments. At the same time, broader calls for 'explainable AI', rules of evidence for AI, and judicial education need to be heeded to ensure that any COR regime enacted by an armed force is capable of being dealt with properly and justifiably.[109]

[108] Wang (n 105).

[109] Katherine Quezada-Tavárez, Plixavra Vogiatzoglou, and Sofie Royer, 'Legal Challenges in Bringing AI Evidence to the Criminal Courtroom' (2021) 12 New Journal of European Criminal Law 531.

Cambridge Law Review (2023) Vol 8, Issue 1, 25–40

Does the United Kingdom Need the 'Brexit Freedoms Bill'?

AMY APPENTENG DANIELS*

ABSTRACT

On 25 October 2022, the 'Brexit Freedoms Bill' was given its second reading in the House of Commons. Here, it was described as 'the culmination of the Government's work to untangle the United Kingdom from nearly 50 years of EU membership'. The ideological stance accompanied an earlier practical justification for its introduction. In fact, Lord Frost (2021 Cabinet Office minister for the Brexit Opportunities Unit) had previously described the necessity of removing the special status of 'retained EU law' ('REUL'), a category of UK laws encompassing legislation, case law and EU principles. He defined these as having 'intrinsically less democratic legitimacy' than UK-initiated laws. The Brexit Freedoms Bill thus provides a sunset clause to facilitate the automatic expiry of REUL on 31 December 2023, unless these are salvaged by ministers. The Bill's two processes of adopting a sunset mechanism and facilitating ministerial reform (rather than the conventional route of legislative reform done by Parliament) have stirred debates over the appropriateness of the Bill as it currently stands. More broadly, a question should be asked of whether the Brexit Freedoms Bill represents a *will* rather than a *need* for reform, justifying its contested measures. This inquiry is brought into stark relief by the selective exclusion of the financial services sector from the remit of the Bill. The resolution of such matter will determine whether the controversial provisions should be tolerated as necessary to deliver an exigency for reform, or whether they are to be challenged further to achieve a more suitable bill.

Keywords: Brexit, EU Law, EU principles, retained EU law, financial services, MiFID

I. INTRODUCTION

On 22 September 2022, (former) Business Secretary, Jacob Rees-Mogg, introduced the Retained EU Law (Revocation and Reform) Bill[1]—known as the 'Brexit

* LLM Candidate (LSE), LLB (SOAS). I am grateful to Dr Maurizio Ghirga (Bank of Italy) for his helpful guidance and feedback. All errors remain mine.

[1] Retained EU Law (Revocation and Reform) HC Bill (2022–23) [156] ('Brexit Freedoms Bill').

Freedoms Bill'—to the House of Commons, where it was given its first reading. A second reading, where MPs had the opportunity to debate the general principles and themes of the Bill, took place on 25 October 2022. Here, Business Under-Secretary, Dean Russell, speaking in lieu of Rees-Mogg, referred to the Brexit Freedoms Bill as 'the culmination of the Government's work to untangle the United Kingdom from nearly 50 years of EU membership'.[2] The reasoning provided has the distinctive ideological overtone of UK emancipation, to be sublimated through legal sovereignty. Yet, although ideology might represent a necessary condition for the introduction of the Bill, it is an insufficient condition per se. A more relevant criterion for its introduction should be whether the changes contemplated by the Bill meet a real reforming requirement in the UK statute book.

The Brexit Freedoms Bill's intention (as stated in the Explanatory Notes accompanying the Bill) is to 'provide the Government with all the required provisions that allow for the amendment of retained EU law (REUL) and remove the special features it has in the UK legal system'.[3] In assessing this objective, one should focus on both the 'special features' that REUL is said to possess in the UK statute book; and on the effects of granting 'all the required provisions' that would create a suite of powers for ministers either to revoke or assimilate REUL. Defining REUL contextualises the changes the Bill is seeking to create: its details will thus be discussed below.

The concurrent focus on REUL's 'special features' and on the provisions empowering ministers will firstly provide insight into the legitimacy of the measures in the Bill; and secondly, it will highlight whether the Bill is merely enforcing an ideological ambition for change or whether it is implementing necessary legal reform.

Indeed, a tension appears to be at play when considering the Bill's purpose. On the face of it, Brexit requires a reform of the UK statute book to purge it of redundant EU references and principles. Yet, this article suggests that, upon reflection, the tactical exclusion of a pivotal sector (that of financial services) from the reforms contemplated in the Bill reveals that the proposed changes are more of a 'will' than a 'need'. In other words, swift reforms are presented as a necessity until they touch on an inconvenient sphere.

Desire for vis-à-vis necessity of reform will impact the appropriateness of the Bill as it currently stands, especially given its time-sensitive sunset clause, surveyed below. In fact, the more urgent the Bill is, the higher the level of tolerance one can have towards divisive measures it might contain. That is, if reforming the UK statute book is truly a pressing issue, then the Bill's controversial measures could be reframed as meeting a genuine exigency. Conversely, if the Bill merely seeks to buttress the Brexit ideology, then its contested processes will need to be

[2] HC Deb 25 October 2022, vol 721, col 183.
[3] Explanatory Notes to the Retained EU Law (Revocation and Reform) Bill, para 1.

recast in the public arena to assess whether possible damaging provisions need to be rejected.

Following this introduction, five main sections advance the discussion. Section II traces the background to REUL's existence as a category of laws, considering its hybrid position as both reflective of the EU legal order it crystallises and of the legislative emancipation of a post-Brexit UK. A breakdown of the subcategories of REUL is further offered in this section, highlighting the complexity of this body of laws. Section III then examines how the Bill nullifies the principle of supremacy of EU law through new interpretative principles foregrounding domestic laws. From there, Section IV critically analyses the mechanics of the Brexit Freedoms Bill, firstly by reviewing the operation of its sunset clause and contrasting this with a previous use of this expiry technique; and secondly, by discussing the controversy behind ministerial powers. A scrutiny of the Bill's impact on the financial services sector is then presented in Section V, with a brief overview of potential reverberations in the sensitive sectors of data protection, employment, and environmental law, as highlighted by the Public Law Project. Finally, concluding thoughts are offered in Section VI, with remarks on how a change in premiership has not led to a deprioritisation of the Bill, and with assessments of solutions to the two problems of vagueness and parliamentary scrutiny surrounding the Bill.

II. Retained EU Law

A. A watershed date

To explain how REUL as a category of laws came to be, a key date—31 December 2020 (referred to as 'Implementation period (IP) completion day')[4]—has to be borne in mind. This marked the end of transitional arrangements arising from UK-EU Brexit negotiations.[5] Such negotiations had determined that, despite not being a member of the EU's political institutions (and thus having no voting rights), the UK would still be subject to EU rules and remain part of the single market and of customs union until the IP completion day.[6]

Following the IP completion day, the UK opted to continue to provide legal continuity and certainty to businesses and individuals by ensuring a gradual progression between the pre-Brexit legal order and what was to come. This resulted in a decision to take a 'snapshot' of all EU legislation on IP completion day, carrying the legislation over into the UK statute book and rebranding it as 'retained EU

[4] European Union (Withdrawal Agreement) Act 2020 ('EUWAA 2020') s 39(1).

[5] LexisNexis Family Expert, 'Implementation Period (IP) Completion Day Definition' (*LexisNexis*) <www.lexisnexis.co.uk/legal/glossary/implementation-period-ip-completion-day> accessed 21 November 2022.

[6] Georgina Wright and Haydon Etherington, 'Brexit Transition Period' (*Institute for Government*, 2 December 2020) <www.instituteforgovernment.org.uk/explainers/brexit-transition-period> accessed 21 November 2022.

law' or 'REUL'.[7] This category of laws, currently in force, occupies a hybrid position.[8] On the one hand, REUL reflects the EU order from which it derives, as it maintains some EU law principles (with examples of general EU principles being supremacy of EU law, legal certainty, proportionality, equal treatment and subsidiarity).[9] On the other, REUL is a foretaste of the legislative freedom that Brexit will accord the UK once it has achieved full autonomy from EU laws. The results of the latter dimension of REUL are twofold. Firstly, legislative changes made at EU-level after IP completion day are not reflected in REUL,[10] meaning that REUL does not dynamically change with EU changes. Secondly, post-IP completion day, Parliament can pass domestic legislation to remove any undesired effect of EU legislation.[11] Yet, this means effecting change at a slow pace, in accordance with law-making timelines.[12] Besides that, in the two years preceding IP completion day, over 600 pieces of UK secondary legislation made around 80,000[13] amendments to REUL.[14] Such amendments, however, were mostly of a technical nature: their function was to ensure the clarity and operability of laws that would apply 'purely in a UK domestic context'.[15]

In summary, REUL maintains a special status in its liminality: it ensures legal continuity (and consequently business certainty) by retaining ties to EU law, whilst offering a glimpse into the legislative independence Brexit offers.

B. Turning to numbers and subtypes

The REUL catalogue reported by the UK Government initially counted 2,417 pieces of legislation spanning across 21 sectors of the UK economy, with the top three being Agriculture, Forestry and Fishing; Transportation and Storage;

[7] Catherine Barnard, 'Commentary: REUL (Retained EU Law) And Lord Frost' (*UK in A Changing Europe*, 17 December 2021) <https://ukandeu.ac.uk/reul-lord-frost/> accessed 21 November 2022.

[8] Catherine Barnard, 'Retained EU Law in the UK Legal Orders: Continuity Between the Old and the New' (2021) University of Cambridge Faculty of Law Research Paper 27/2021 <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3947215> accessed 21 November 2022.

[9] 'General principles of EU law' (*Thomson Reuters Practical Law*) <https://uk.practicallaw.thomsonreuters.com/w-018-9132?contextData=(sc.Default)&transitionType=Default&firstPage=true> accessed 21 November 2022.

[10] David Thorneloe, 'Retained EU Law in the UK after Brexit' (*Pinsent Masons*, 5 January 2021) <www.pinsentmasons.com/out-law/guides/retained-eu-law-uk-after-brexit> accessed 21 November 2022.

[11] Barnard (n 8) 1.

[12] ibid.

[13] This figure varies: the Cabinet Office reports in a Memorandum to the Delegated Powers and Regulatory Reform Committee that 'more than 1,000 statutory instruments [were] made by the UK and devolved governments', estimating 'more than 100,000 amendments' to REUL. Cabinet Office, 'Retained EU Law (Revocation and Reform) Bill: Memorandum from the Cabinet Office to The Delegated Powers and Regulatory Reform Committee' (*UK Parliament*, 22 September 2022) <https://bills.parliament.uk/bills/3340/publications#collapse-publication-delegated-powers-memorandum> accessed 21 November 2022.

[14] Thorneloe (n 10).

[15] ibid.

and Financial and Insurance Activities (Table II.1).[16] On 8 November 2022 a further 1,400 pieces of REUL emerged from the research of The National Archives.[17] The latest calculation by the Government on 30 January 2023 brings the total count to over 3,700 pieces of legislation, concentrated over 400 unique policy areas.[18] This figure is to be updated on a quarterly basis, with government departments working to identify further REUL.[19]

TABLE II.1

Top Sectors Retained EU Law Fall Under	
Sector	*Number of REUL*
Agriculture, Forestry and Fishing	493
Transportation and Storage	482
Financial and Insurance Activities	365
Manufacturing	347
Public Administration and Defence; Compulsory Social Security	133

Source: Government Reporting

Although this figure provides an overall picture of the scale of reform the Brexit Freedoms Bill is concerned with, it offers only a narrow view of what the REUL category actually includes. A more suitable overview is offered by understanding REUL as a heading nesting three types of laws: (a) REUL *stricto sensu*; (b) retained EU case law; and (c) general REUL principles.[20]

(i) REUL Stricto Sensu

REUL *stricto sensu* is simply legislation: both EU legislation incorporated into UK law before IP completion day ('EU *derived* or *preserved* legislation');[21] and EU legislation directly applicable or directly effective but which had not been

[16] Cabinet Office, 'Retained EU Law – Public Dashboard' (*Tableau Public*) <https://public.tableau.com/app/profile/governmentreporting/viz/UKGovernment-RetainedEULawDashboard/Guidance> accessed 9 November 2022.

[17] George Parker, 'UK Plan to Scrap All EU Laws Suffers New Setback' *Financial Times* (London, 8 November 2022) <www.ft.com/content/0c0593a3-19f1-45fe-aad1-2ed25e30b5f8> accessed 9 November 2022.

[18] Department for Business and Trade, Department for Business, Energy & Industrial Strategy and Grant Shapps, 'Research and Analysis: Retained EU Law Dashboard' (*Gov.uk*, 22 June 2022, last updated 30 January 2023) <www.gov.uk/government/publications/retained-eu-law-dashboard> accessed 1 February 2023.

[19] ibid.

[20] Barnard (n 8) 3.

[21] European Union (Withdrawal) Act 2018 ('EUWA 2018') s 2.

specifically implemented into UK law before IP completion day ('EU *converted* legislation').[22]

(ii) Retained EU Case Law and General REUL Principles

Post-IP completion day, UK courts and tribunals are no longer bound by principles or decisions from the European Court of Justice ('ECJ'), nor can they refer matters to it.[23] In line with the notion of legal continuity that underlies REUL, European decisions pre-IP completion day—together with the EU's interpretative methods—are also retained (unless modified after IP completion day). Additionally, an extension is made for the Court of Appeal to depart from retained EU case law (a power previously granted only to the UK's highest court, the Supreme Court).[24] When presented with the opportunity, the Court of Appeal, however, refused to do so, noting that this power was to be exercised cautiously.[25] This demonstrated adherence to EU-level decisions and deference to their influence over the UK context: further evidence of REUL being perceived as tied to the pre-Brexit order.

C. Replacing the placeholder

The three categories, only presented in headline form above,[26] might give some indication of the complexity of REUL. Complexity does not, however, mean finality: REUL is only a placeholder, only part of a process towards the restoration of the UK's legislative sovereignty. In the words of Lord Frost, finalising this process would mean 'to remove the special status of retained EU law so that it is no longer a distinct category of UK domestic law, but normalised within [the UK's] law, with a clear legislative status'.[27]

Lord Frost was the 2021 Cabinet Office minister for the Brexit Opportunities Unit, succeeded in 2022 by (former) Business Secretary, Rees-Mogg (who is now a backbencher under Rishi Sunak's premiership). Lord Frost explained that the rationale for the intended overhaul of REUL was that 'laws agreed elsewhere have intrinsically less democratic legitimacy than laws initiated by the Government of this country'.[28] Besides the birthplace of laws, the influence of EU principles, still present through REUL as illustrated above, have compounded the desire to

[22] Barnard (n 8) 7–15. Barnard provides EU Directives as examples of 'EU derived or preserved legislation' and EU Treaty provisions, EU Regulations, and EU Decision as examples of 'EU converted legislation'.
[23] EUWA 2018 (n 21) s 6(1).
[24] EU (Withdrawal) Act 2018 (Relevant Court) (Retained EU Case Law) Regulations 2020, SI 2020/1525.
[25] *TuneIn Inc v Warner Music UK Ltd* [2021] EWCA Civ 441, [2021] Bus LR 1119 [73]–[89] (Arnold LJ).
[26] For a more detailed analysis of the subtypes of REUL, see Barnard (n 8).
[27] Lord Frost, 'Lord Frost Statement to The House of Lords: 16 September 2021' (House of Lords, London, 16 September 2021) <www.gov.uk/government/speeches/lord-frost-statement-to-the-house-of-lords-16-september-2021> accessed 21 November 2022.
[28] ibid.

reform REUL, in a post-Brexit UK.[29] Epitomising such influence is the so called 'Marleasing principle',[30] which established that courts of EU member states have a duty to interpret national law in a way that gives effect to EU law. Therefore, reforming REUL is seen as a step in the direction of recovering the separate identity of UK law.

With this vision, the Brexit Freedoms Bill aims to operate on two fronts, firstly by nullifying the principle of supremacy of EU law and other interpretative principles of EU law, and secondly by imposing a sunset deadline to the revocation of REUL within which ministers are empowered to restate REUL as UK law or allow the lapsing of REUL. These two aspects will be examined in turn.

III. Supremacy of Domestic Law

A significant consequence of the Bill passing into law would be the abolition of the supremacy of EU law[31] and of general principles of EU law.[32]

Clause 4(1)(A2) of the Brexit Freedoms Bill provides a new interpretative instruction: that REUL provisions be read and implemented in a way that is compatible with domestic enactments. This is re-emphasised in clause 4(1)(A2)(b) which states that any provision of REUL that is incompatible with domestic laws is subject to such domestic laws.

This clause of the Brexit Freedoms Bill is a clear departure from REUL's distinction between pre and post IP completion day. Now the message becomes that irrespective of when a law was passed, it need not show deference to EU law.

A similar reasoning is applied to general EU principles and ECJ judgments. EU principles no longer affect the interpretation of the UK statute book, and even the name 'REUL' is changed to 'assimilated law' to remove references to the EU source. With regards to the development of domestic case law, the Bill enables UK courts' divergence from retained EU case law and allows them to go one step further by making an 'incompatibility order'[33] in case of discrepancy between REUL and any domestic enactment.[34]

All these changes are time-bound, with an imminent deadline provided through the operation of a sunset clause. In the set timeframe, courts are not the

[29] Graeme Cowie and Ali Shalchi, 'Research Briefing: Retained EU Law (Revocation and Reform) Bill 2022-23' (*House of Commons Library*, 17 October 2022) <https://commonslibrary.parliament.uk/research-briefings/cbp-9638/> accessed 21 November 2022.

[30] Case C-106/89 *Marleasing SA v La Comercial Internacional de Alimentación SA* [1990] ECR I-4135.

[31] Brexit Freedoms Bill (n 1) cl 4.

[32] ibid cl 5.

[33] The Bill does not offer thorough details of how the incompatibility order would work. Possibly, this could be similar to a 'declaration of incompatibility', a current UK constitutional law feature that enables UK judges to declare that a statute is incompatible with the European Convention of Human Rights. This is a simple declaration that acts as a signal to Parliament to eventually intervene to remove the incompatibility.

[34] Brexit Freedoms Bill (n 1) cl 9.

only domestic institution to be empowered: ministers are authorised to revoke, reaffirm or replace REUL. This is reviewed below.

IV. Sunset Clause and Ministerial Powers

A. Sunset Clause

The revocation of REUL will take place through a sunset clause (clause 1 of the Brexit Freedoms Bill) capturing both EU-derived subordinate legislation[35] and retained direct EU legislation[36]. The automatic expiry of REUL will take effect on 31 December 2023 unless the retained EU laws are preserved by ministers.[37] The Explanatory Notes of the Bill posit that the logic behind the use of a sunset clause is on the one hand, the acceleration of reform to the benefit of businesses and consumers; and on the other, the increase of business certainty about when a 'new domestic statute book' will come into effect.[38] This formal justification for the sunset clause-method of legally breaking ties with EU law thus emphasises practical benefits, fortifying the layer of ideological benefits advanced to introduce the Bill as a whole.

The use of a sunset provision is certainly not new to the Government, with the emergency Coronavirus Act 2020 having made use of a two-year sunset clause to impose a time limit on most of ministers' emergency provisions.[39] The Government's ability to extraordinarily exercise its powers through emergency arrangements, however, was tempered by clause 98 of the Act which provided a six-month parliamentary review mechanism. Through this, insofar as it was practicable to do so, ministers had to arrange the debate and vote of their motions in the House of Commons within seven days of the end of each six-month interval.[40] If the motions were rejected by the House of Commons, ministers had 21 days within which to ensure the expiry of the relevant temporary provisions.[41]

Returning to the Brexit Freedoms Bill, a noteworthy point must be raised. Clause 2 of the Bill allows ministers to extend the 31 December 2023 sunset deadline to a time no later than 23 June 2026 (the decennary of the Brexit Referendum). A further parallelism can be drawn with the Coronavirus Act 2020: this Act reined in ministerial powers, as intermediate scrutiny by the House of Commons effectively reduced the arc of a two-year sunset period into six-month blocks. Conversely, the expansionary provision in clause 2 of the Brexit Freedoms Bill achieves the opposite of constraining ministerial powers: it lengthens the

[35] Brexit Freedoms Bill (n 1) cl 1(1)(a).
[36] ibid cl 1(1)(b).
[37] ibid cl 1(2).
[38] Explanatory Notes (n 3) para 18.
[39] Coronavirus Act 2020, s 89(1).
[40] ibid s 98(3).
[41] ibid s 98(1).

exercise of such powers from an initial one-year period to a four-year period, without specifying how Parliament will oversee the use of such powers.

Therefore, the possible deadline extension appears to exacerbate what might already be considered a democratic deficit in the Brexit Freedoms Bill. In fact, according to the prevalent opinion of witnesses at the Committee stage of the Bill,[42] the ministerial power to amend REUL undermines parliamentary sovereignty, the paramount UK constitutional principle[43] that holds Parliament to be the supreme law-making authority in the country. This point is reemphasised in the third reading of the Bill, where Michael Amesbury, an opposition Labour Party MP, notes that parliamentary sovereignty consists of giving control to Members of Parliament rather than the Executive or Whitehall bureaucrats.[44]

This is not, however, a universally accepted interpretation of the principle. A minority opinion advanced by Sir Stephen Laws KC[45] (First Parliamentary Counsel[46] from 2006 to 2012) holds that the tenet of parliamentary sovereignty is a 'myth' and rather defines Parliament as a 'political filter for legislation'. In light of this, Government can still be made accountable to Parliament through the latter's scrutiny of the 'politically salient' aspects of legislation.[47] According to this view, Parliamentary oversight of the 'mainly technical',[48] bureaucratic task of removing 'legally inoperable' EU legislation would be an unnecessary complexity that contravenes good governance. [49]

The issue then becomes whether the Bill itself should better define the limits of mere ministerial review by indicating the cases in which Members of Parliament should be allowed to step in, and consult on or challenge political aspects of critical importance to the electorate.

B. Clause 15 ministerial powers

The crux of the revocation and reform powers contained in the Bill can be found in clause 15, a reading of which raises the question of the extent to which such powers are fundamentally an 'executive power-grab'.[50]

[42] PBC Deb (Bill 156) 8 November 2022, cols 15–16, 28, 30–31, 34.
[43] A principle emphasised in EUWAA 2020 (n 4) s 38.
[44] PBC Deb (Bill 156) 18 January 2023, col 398.
[45] Current Senior Research Fellow at Policy Exchange, a British conservative think tank based in London; Sir Stephen Laws KC appeared as the first witness at the PBC Deb (Bill 156) 8 November 2022, see col 5.
[46] First Parliamentary Counsel is the head of the Office of Parliamentary Counsel, a group of government lawyers who draft government legislation introduced to Parliament.
[47] PBC Deb (Bill 156) 8 November 2022, cols 5–6.
[48] PBC Deb (Bill 156) 18 January 2023, col 433.
[49] ibid col 397.
[50] Lord Anderson of Ipswich, 'Retained EU Law (Revocation and Reform) Bill: Notes for Remarks' (Speech before the Bar European Group, Matrix Chambers, London, 19 October 2022) para 4 <www.daqc.co.uk/wp-content/uploads/sites/22/2022/10/RETAINED-EU-LAW.pdf> accessed 7 November 2022.

According to clause 15, a minister can either revoke 'secondary' REUL without replacement[51] (causing a legislative gap); or replace the revoked law with a provision the minister considers appropriate and with[52] or without[53] the requirement that it have 'the same or similar objectives' to the replaced law. This clause remains ambiguous in its failure to articulate when the replacement option must pursue similar objectives and when it can ignore this requirement.

Additionally, opacity is found in the terminological choice. As noted by the national legal charity, the Public Law Project (PLP),[54] the term 'secondary' REUL adopted in clause 15 misleadingly suggests that the laws that might be changed are technical in nature, rather than including substantive rights. This is on account of a 'category error' that equates EU secondary legislation (called 'secondary' only to distinguish them at EU level from treaties, which are called 'primary') to UK secondary legislation (which is law known as a statutory instrument created by ministers).[55] Calling the REUL in clause 15 'secondary' is thus a *misnomer* that gives the illusion of ministers operating within their usual mandate of passing statutory instruments.

V. Impact on Financial Services

An exceptional point of the Brexit Freedoms Bill is clause 22(5) which excludes the application of the sunset clause to specific financial services legislation, namely:

1. Anything referred to in Schedule 1 to the Financial Services and Markets Act 2022;[56]
2. Rules made by the Financial Conduct Authority (FCA), the Prudential Regulation Authority (PRA) or the Bank of England (BoE);[57] and
3. Requirements or directions imposed by the Payment Systems Regulator.[58]

An analysis by practitioners in financial regulation services[59] indicates that this exclusion is because of the publication of the Financial Services and Markets

[51] Brexit Freedoms Bill (n 1) cl 15(1).
[52] ibid cl 15(2).
[53] ibid cl 15(3).
[54] Samuel Willis, 'Retained EU Law (Revocation and Reform) Bill Second Reading Briefing: Recommendations on Ensuring Unconstrained Legislative Powers are Not Transferred to Ministers' (*Public Law Project*, 24 October 2022) para 13 <https://publiclawproject.org.uk/resources/retained-eu-law-revocation-and-reform-bill-second-reading-briefing> accessed 7 November 2022.
[55] ibid.
[56] Brexit Freedoms Bill (n 1) cl 22(5)(a).
[57] ibid cl 22(5)(b).
[58] ibid cl 22(5)(c).
[59] I am grateful to Caroline Dawson (Partner) and Paul Lenihan (Senior Associate) at Clifford Chance for their helpful comments. All errors remain mine.

('FSM') Bill before the Brexit Freedoms Bill.[60] In fact, the FSM Bill contains provisions conferring powers on HM Treasury to make secondary legislation, and on the regulators to make rules, which replace REUL relating to financial services. Including financial services legislation in the scope of the Brexit Freedoms Bill would have been thus, at best redundant and at worst in conflict with a more specific regime. Furthermore, the FSM distinguishes itself from the Brexit Freedoms Bill in a way that makes it more suitable to reform the body of financial services REUL. Firstly, the FSM Bill does not include any sunset mechanism: this avoids ministerial haste in the review process. Secondly, the FSM Bill contains a finite list of laws to be reviewed, specified in the annexed Schedule 1.

The approach under the FSM Bill has been described, in antithesis to that of the Brexit Freedoms Bill, as 'responsible and measured'.[61] Yet, risk is not completely eliminated from the FSM Bill, as it also provides a sweep-up provision enabling the repeal of all EU-derived legislation (excluding primary legislation) relating to financial services which is not captured in Schedule 1.[62] This signals that even purportedly well-drafted future legislation (such as the FSM Bill) cannot guarantee a total identification of all REUL to be reviewed. Perhaps this note could be a useful reminder to adopt a more clement perspective on the Brexit Freedoms Bill and focus on what this Bill could ultimately achieve. To this end, appeals have been made[63] to use the reforming objective of the Brexit Freedoms Bill, and specifically its provisions displacing EU law supremacy with UK law supremacy, to correct the so-called 'MiFID override'.

A. MiFID OVERRIDE

The Markets in Financial Instruments Directive ('MiFID') override is a legislative fix contained in Article 4(4) of the Regulated Activities Order 2001 ('RAO')[64] that creates the *paradox* (in a post-Brexit UK) of holding EU financial services regulations to be superior to UK legislation in case of discrepancy; and the *inconvenience* for firms and investors having to review constantly both the RAO and definitions and exemptions in MiFID II.

The RAO sets out what activities and instruments are regulated in the UK, covering the same subject matter of the EU equivalent investment-business

[60] The FSM Bill had its first reading on 20 July 2022, whilst the Brexit Freedoms Bill had its first reading on 22 September 2022.

[61] Words by Mark Fenhalls, (then) Chair of the Bar Council and witness at the PBC Deb (Bill 156) 8 November 2022, col 28.

[62] Financial Services and Markets HC Bill (2022-2023) [146] ('FSM Bill') cl 1(5).

[63] Barnabas Reynolds, Thomas Donegan and Sandy Collins, 'The Brexit Freedoms Bill and the MiFID Override for Financial Services Regulation' (*Shearman & Sterling*, 24 October 2022) <www.shearman.com/en/perspectives/2022/10/the-brexit-freedoms-bill-and-the-mifid-override-for-financial-services-regulation#page=1> accessed 7 November 2022.

[64] A statutory instrument made under the Financial Services and Markets Act 2000.

directives, MiFID I[65] and MiFID II[66]. Yet, there is not perfect congruity between the RAO and the MiFID framework, specifically with respects to regulatory exemptions. This issue has been addressed by allowing the two frameworks to coexist in the UK and requiring, under the MiFID override, a concurrent look at the frameworks and a prevailing of the narrower MiFID exemptions over the broader RAO exemptions. Practitioners have confirmed the inconvenience of this parallel scrutiny,[67] drawing from their experience of advising clients moving from an existing unregulated business in the UK to an activity touching on the regulated sphere. In these cases, such clients would need to know the extent to which they can rely on exemptions to prove that they do not fall under the regulated market. A double reference text increases the complexity of ascertaining this.

Such call from the world of practice therefore holds that the Brexit Freedoms Bill could be used in its explicit reneging of EU authority to incentivise the creation of a single UK source, clearly defining regulations and exemptions. This assumes that some financial services areas will escape the list in the FSM Bill, and thus inevitably fall under the purview of the Brexit Freedoms Bill (unless explicitly exempt under it).

B. Pragmatism prevailing?

Besides some practitioners' desire to see a direct impact of the Bill in the MiFID area, wider questions emerge from the pre-emptive exclusion of the financial services sector from the remit of the Brexit Freedoms Bill.

The decision to shield financial services from sunsetting indicates the dominance of pragmatic calculations over the ideological stance of the Bill. That is, so long as business needs favour the application of EU laws, these can remain. At the same time, this position then betrays the actual necessity of reform of the UK statute book as a whole. The notion of disapplying EU laws as an urgent priority underscored the idea of accelerating reform through a sunset clause. Yet, this was easily cast aside when it proved to be inexpedient. One might conclude that the reform justification of the Brexit Freedoms Bill veils a desire to advance the Brexit political outcome to the next legal stage. The application of a sunset over other sectors is however still of consequence to the business world: this is briefly addressed next.

[65] Directive 2004/39/EC of the European Parliament and of the Council of 21 April 2004 on markets in financial instruments amending Council Directives 85/611/EEC and 93/6/EEC and Directive 2000/12/EC of the European Parliament and of the Council and repealing Council Directive 93/22/EEC [2004] OJ L145/1.

[66] Directive 2014/65/EU of the European Parliament and of the Council of 15 May 2014 on markets in financial instruments and amending Directive 2002/92/EC and Directive 2011/61/EU [2014] OJ L173/349.

[67] See n 59.

C. Impact on other sectors

Although the exclusion of financial services from the purview of the Brexit Freedoms Bill might ensure continuity for businesses planning future activities, the reality is that businesses are affected by legal changes in other economic sectors.

The PLP's compilation of the scope of rights and protections currently guaranteed under REUL in the areas of data protection, employment and environmental law (Table V.1) reveals the breadth of these sensitive fields.[68] One can immediately comprehend the good business sense that comes from proper information handling under the General Data Protection Regulation's ('GDPR')[69] extensive regime of principles, rights and obligations. Similarly, businesses with employees will have to confront the intricate web of labour law protections afforded to such employees. These cover matters ranging from maximum working hours, length of night work and annual leave entitlements;[70] to the protection of employees from dismissal for mere transfer reasons;[71] or the favourable treatment of fixed-term[72] or agency workers[73]. Moving then to environmental considerations: businesses providing development projects[74] or public plans[75] are currently subject to assessments of their environmental impact, demonstrating the deemed necessity of external legal requirements to further environment protection objectives.

Therefore, changes to the three areas of data protection, employment, and environmental law may also benefit from a pondered review, a type of deliberation that might only occur when not rushing towards an imminent deadline.

The issue with sunsetting legislation in these fields is not only to be seen in the legal uncertainty that might loom over businesses, but in the behavioural disincentives that legislative voids (even if the voids are only short-lived) risk creating. On this point, Professor Catherine Barnard[76] observed at the Committee stage of the Brexit Freedoms Bill that, in the absence of legislation mandating prescribed actions, businesses seeking to cut costs will not necessarily comply with high standards.

[68] Willis (n 54).

[69] 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.

[70] Working Time Regulations 1998, SI 198/1833.

[71] Transfer of Undertakings (Protection of Employment) Regulations 2006, SI 2006/246.

[72] Fixed-Term Employees (Prevention of Less Favourable Treatment) Regulations 2002, SI 2002/2034.

[73] Agency Workers Regulations 2010, SI 2010/93.

[74] Directive 2011/92/EU of the European Parliament and of the Council of 13 December 2011 on the assessment of the effects of certain public and private projects on the environment [2011] OJ L26/1.

[75] Directive 2001/42/EC of the European Parliament and of the Council of 27 June 2001 on the assessment of the effects of certain plans and programmes on the environment [2001] OJ L197/30.

[76] Professor of European and Employment Law at the University of Cambridge and witness at the PBC Deb (Bill 156) 8 November 2022, see col 11.

TABLE V.1

Examples of Rights and Protections Potentially Affected by Augmented Powers		
Sector	*REUL*	*Description*
Data Protection	General Data Protection Regulation (GDPR)	Source of important data protection rights (for instance, right to be informed, right of access, right to rectification, right to erasure)
Employment	Working Time Regulations 1998 (SI 198/1833)	Maximum weekly working time and right to holiday pay (including case law on formula for calculating holiday pay)
	Transfer of Undertakings (Protection of Employment) Regulations 2006 (SI 2006/246) (TUPE)	Protects the rights of workers whose jobs are outsourced or transferred to another business
	Part-Time Workers (Prevention of Less Favourable Treatment) Regulations 2000 (SI 2000/1551)	Protects part-time workers from being treated less favourably than full-time workers just because they are part-time
	Information and Consultation of Employees Regulations 2004 (SI 2004/3426)	Requires employers to establish arrangements for informing and consulting their employees
	Health and Safety (Consultation with Employees) Regulations 1996 (SI 1996/1513)	Employers have a duty to consult their employees, or their representatives, on health and safety matters
	Fixed-Term Employees (Prevention of Less Favourable Treatment) Regulations 2002 (SI 2002/2034)	Protects fixed-term workers from being treated less favourably than full-time workers just because they are fixed-term
	Agency Workers Regulations 2010 (SI 2010/93)	Agency workers are entitled to the same or no less favourable treatment for basic employment/working conditions
Environment	Habitats Directive 92/43/EEC (and implementing regulations)	Protects special habitats and/or species (such as, through the designation of Special Areas of Conservation)
	Environmental Impact Assessment Directive 2011/92/EU (and implementing regulations)	Development projects that are likely to have a significant environmental impact must be identified and have their environmental impact assessed
	Strategic Environmental Assessment Directive 2001/42/EU (and implementing regulations)	Public plans and projects are subject to an assessment of environmental impact

Source: The Public Law Project

VI. CONCLUSIONS

Looking back to 2016 when the Brexit Referendum (that is, the trigger of the UK's divorce from the EU) took place, one cannot help but reflect on the UK having had six years to consider what laws to retain and what laws to do away with. Still,

we must be reminded that these six years have been far from ordinary:[77] five premierships, two monarchies and two global crises. This is the background to the Brexit Freedoms Bill's scrutiny.

A. Recent political developments

The most recent change in premiership might be of consequence to the Bill's future. On 20 October 2022, prime minister Liz Truss, successor to Boris Johnson in heading the Conservative Party, announced her resignation a mere 44 days after taking office. The vacant role was occupied on 25 October 2022, by Rishi Sunak, Chancellor of the Exchequer in Boris Johnson's cabinet who had resigned from his role following the Pincher scandal. Sunak's new cabinet featured a replacement of Rees-Mogg with Grant Shapps as Business Secretary, leading to speculations of a possible deprioritisation of the Bill. Yet, Sunak himself had been a promoter of a systematic review of REUL during his campaign in the Conservative Leadership contest against Truss.[78] With reference to REUL, his promotional message indicated that he would review 'all 2400 of them' within the first 100 days of his premiership.[79] With the recent resurfacing of additional laws that bring the current total to over 3,700, the concern of whether there is sufficient capacity in governmental departments to conduct the review becomes of true relevance. This might be the most pragmatic consideration on which to reflect; it is not, however, the most severe feature that has attracted criticisms and suggestions of corrective measures.

B. Possible solutions

Two main problems can be identified at the centre of major denunciations of the Bill. First, the indefinite number of REUL that could be the subject of reform. The Government's new exercise of updating its catalogue of REUL on a quarterly basis offers only a partial solution: a periodic update in fact is simply an 'open-ended expansion of the list of EU laws'[80] that fails to correct the present uncertainty. Such vagueness will likely prove uninviting for entities who might otherwise be interested in conducting business in the UK. Moreover, if further unanticipated REUL were to emerge between now and the December 2023 deadline (a scenario that is predicted by the Government)[81], these would burden the

[77] Alison Young, Sir David Williams Professor of Public Law at the University of Cambridge and witness at the PBC Deb (Bill 156) 8 November 2022, see col 12.
[78] Rishi Sunak (*Twitter*, 8 August 2022) <https://twitter.com/RishiSunak/status/1556590394170818560?s=20&t=2sh3l43Fc6NgeK7qQeSxiA> accessed 30 November 2022.
[79] ibid.
[80] Peter Foster and George Parker, 'UK Review of EU Laws Expanded After 1,000 Pieces of Legislation Added' *Financial Times* (London, 30 January 2023) <www.ft.com/content/060b957b-97e8-4580-ad48-a538fcc423fd> accessed 1 February 2023.
[81] See n 18.

load and planning of an already encumbered civil service. A solution might be to set a deadline to the sunset only once a definitive list of the legislation to be modified has been conclusively identified. Thus, replication of the FSM Bill's template has been welcomed by consultees in Parliament,[82] with the added suggestion of imposing different sunset deadlines on regulations of differing magnitude and urgency.[83] On this point, Mark Fenhalls KC[84] further indicates that a list would be the basis for a 'proper ministerial division of responsibility as to who is doing what'.[85] As the Bill is currently proceeding in the House of Lords, such marshalled list of REUL could be requested by the Lords themselves.

The second problem is the contended lack of sufficient democratic input by Parliament. The mechanics of how parliamentary scrutiny should work are a glaring omission in the Bill. Sir Stephen Laws has indicated that provisions about parliamentary procedure need not be set in legislation;[86] however, he fails to offer a solution as to where these should be placed. A more tenable course of action would be that proposed by the Bar Council: requiring a consultation and allowing sufficient time for Parliament to debate any REUL that is restated or revoked.[87] It appears that this endeavour would imply that questions of timing and sunsetting addressed as the first major problem would be coming full circle.

Overall, the Brexit Freedoms Bill would indeed represent the legal crowning of the Brexit vote, as it aims to expedite the independence of the UK's statute book from the influence of EU laws. The selective exclusion of the financial services sector from the sunsetting mechanism contemplated in the Bill, however, is revelatory of a strategic disdain for EU-initiated laws. Despite the Bill being presented as a necessary step, this single act might downgrade its urgency, leading one to suggest that wider consensus be reached before it becomes an Act in its current form.

[82] PBC Deb (Bill 156) 8 November 2022, cols 12; 32–34.
[83] ibid col 12.
[84] See n 61.
[85] PBC Deb (Bill 156) 8 November 2022, cols 32–34.
[86] Text to n 47.
[87] Text to n 50.

Cambridge Law Review (2023) Vol 8, Issue 1, 41–71

Towards an Idea of Digital Asset Ownership

KAN JIE MARCUS HO*

ABSTRACT

As traditional notions of property come into contact with nascent forms of digital assets, courts have questioned whether Fry LJ's seminal statement regarding the lack of a *tertium quid* between chooses in action or possession ought to continue to hold true in modern property law. This article argues to the contrary and contends instead for a third category of property to be developed. In doing so, it draws inspiration from the Law Commission's 2022 Consultation paper, and its proposed third category of property, 'Data Objects', and suggests several tweaks to the Law Commission's model. In Section II, this article argues that the proposal is myopic in some aspects, particularly in scope and associated remedies, and offers solutions to remedy this. In building on the Law Commission's proposal, Section III then offers a comparative study of how common law jurisdictions have treated digital assets and applies these lessons to show the weakness of *Ainsworth*, solidifying the case for a third category to be created. In Section IV, this paper returns full circle to Section II, proposing a reworked third category from the Law Commission's model, which is underpinned by a test based on the types of types of data concerned. This article suggests various entry points forward, and concludes that the effect of developing such a category, and consequently away from *Ainsworth*, will ground property law firmly back within the Hohfeldian 'bundle of rights' model, hence bringing the law back in line with policy and reality.

Keywords: technology Law, property Law, crypto tokens, digital assets, non-fungible tokens

I. INTRODUCTION

It is arguable that the law of property has hit a quandary following the exponential growth in digital assets across all areas of modern society. An appropriate place to set the stage for the problem the law is currently facing—and the problem this paper seeks to address—is the Consultation Paper by the United Kingdom Law Commission in 2022 on Digital Assets.[1] There, the Law Commission noted that the

* BA (Hons) (Law) (*Cantab*) (First Class Honours); LLM (Harvard) (Dean's Scholar Prize Winner); MCIT (UPenn) (Candidate). I am grateful to the anonymous reviewers for their comments on earlier drafts.

[1] Law Commission, *Digital Assets: Consultation Paper* (Law Com No 256, 2022).

English law of property has traditionally recognised only two categories of personal property, these being: (a) things in possession; and (b) things in action.[2] According to the Law Commission, such a bifurcation straddles the advent of digital assets rather uncomfortably, for digital assets 'nevertheless have the characteristics of other objects of property rights'.[3] Hence, the Law Commission suggests a change to the law of property—this being the creation of a third category of property, a category distinct from things in possession and things in action, termed as 'data objects'.[4] In doing so, the Law Commission sets out a set of criteria to determine when a thing would properly fall under the ambit of a 'data object', and applies it to various types of digital assets.

Section II of this paper therefore seeks to evaluate said proposal by the Law Commission, after exploring the basic concepts of the law of property. This paper argues that the current position adopted by the Law Commission remains myopic as to how it applies to other digital assets, particularly given that it focuses far too much on crypto tokens. Indeed, legal uncertainty continues to loom large as it relates to other digital assets, especially in the context of cloud storage and other intangibles. Further, this paper argues that the Law Commission's criteria for 'data objects' could be further reworked, specifically in its definition of 'data', as well as its associated legal remedies. To build on the Law Commission's proposal, Section III of this paper then seeks to explore how common law jurisdictions have treated digital assets in the context of the law of property, through analysing the policy set-up and the juridical technological discourse that has occurred to date. Finally, in Section IV this paper concludes that the Law Commission's proposal, whilst commendable, requires some tweaks. This paper will argue that the Hohfeld's 'bundle of rights' theory serves to inspire the right way forward as to how the law should develop in relation to digital assets.

II. The Law Commission's Proposal

As Soto argues, how legal systems seek to recognise property is essential, for property rights are recognised against the whole world; whilst personal rights are merely recognised against someone who has taken on a relevant legal duty.[5] Indeed, legal property finds itself as the 'indispensable process' that 'fixes and deploys capital', and mankind would be unable to 'convert the fruits of its labour into fungible, liquid forms that can be differentiated, combined, divided, and invested to produce surplus value' if a stable and consistent property framework is lacking.[6] The Law Commission properly noted that the advancement of digital assets would 'exponentially expand the scope of this productive process', as digital assets

[2] ibid para 1.14.
[3] ibid.
[4] ibid para 1.15.
[5] Hernando de Soto, *The Mystery of Capital* (Bantam Press 2000) 164.
[6] ibid.

enhance this process by 'enabling the communication of value via electronic means, which broadens the scope and access to markets and increases the transferability, composability, and liquidity of things of value'.[7] It is therefore important for legal property rights to facilitate said process.

It is trite that the law of property remains better described, rather than defined in a single term. As Edelman posits, the initial problem which faces any analysis of property rights is the 'lack of any coherent definition of property'.[8] Indeed, this quandary is reflected in statute within English law. The Insolvency Act 1986 defines 'property' as "money, goods, things in action, land and every description of property wherever situated and also obligations and every description of interest, whether present or future or vested or contingent, arising out of, or incidental to, property'. [9] In contrast, the definition of 'property' in the Tort (Interference with Goods) Act 1977 explicitly excludes 'things in action and money'.[10] This is justifiable insofar as diverging policy objectives ground the definition within the particular statutes, but it might not be helpful in working out a tenable classification of property in the round.

The Law Commission heralds an escape out of this quandary by endorsing an understanding of the concept of property as 'not a thing at all, but a socially approved power-relationship in respect of socially valued assets, things, or resources'.[11] This 'power relationship' formulation suggests that the legal construct of property consists of three elements: (a) the existence of an asset, thing or resource to which a power or right can relate; (b) the liberty of a person to use the asset, thing, or resource; and (c) the right of a person either to exclude or allow access by another person to that particular asset, thing, or resource. The formulation is underpinned by a relationship between a person and a thing, instead of the notion of a thing in itself. Nevertheless, the Law Commission correctly acknowledges that the logically prior question one must address is what kinds of things exactly can be the subject of a property right. As Professor Birks posits, suitable objects of property are 'the [thing] to which a [property right] relates'.[12] Whether something constitutes a thing, however, is an inherently fuzzy notion, particularly when one is trying to understand where the boundaries of what a 'thing' are.[13] Accordingly, guiding principles have been developed to facilitate effective analysis. The Law Commission observes that five often used criteria which have been developed in this regard: (a) the *Ainsworth* test; (b) that the thing must be rivalrous; (c) excludability; (d) separability; and (e) value.[14] The evaluation of these five

[7] Law Commission (n 1) para 1.5.
[8] James Edelman, 'Property Rights to Our Bodies and Their Products' (2015) 39 University of Western Australia Law Review 47, 52.
[9] Insolvency Act 1986, s 486.
[10] Torts (Interference with Goods) Act 1977, s 14(1).
[11] Law Commission (n 1) paras 2.10, 2.16.
[12] Peter Birks, *An Introduction to the Law of Restitution* (Clarendon Press 1985) 49.
[13] Law Commission (n 1) para 2.28.
[14] ibid para 2.36.

criteria will inform the analysis that follows in relation to both critiques the Law Commission's proposal and the possible ways forward.

A. THE *AINSWORTH* TEST

In *National Provincial Bank v Ainsworth*, Lord Wilberforce set out four characteristics which describe a 'thing' that constitutes 'property', these being that the thing must be 'definable', 'identifiable by third parties', 'capable in its nature of assumption by third parties', and 'have some degree of permanence or stability'.[15] Cutts construes the *Ainsworth* characteristics as somewhat of a 'negative threshold' test for considering when something might attract property rights.[16] In other words, a "thing" that does not fulfil the four characteristics would likely not be considered as attracting property rights. Notwithstanding this, it does not also necessarily follow that a thing will attract property rights just by fulfilling the *Ainsworth* criteria.[17] The *Ainsworth* criteria will be further explored in Section III of this article when discussing common law jurisprudence, but a preliminary comment might be made that this criterion, although helpful as a starting point, does not pull much weight when plunged into the murky depths of edge cases as might be common in digital assets.

B. RIVALROUS

Michels and Millard,[18] alongside several notable scholars, have argued that the concept of rivalrous is a core trait of things which attract property rights. A rivalrous product, as is often found in discourse relating to economics, is something whose 'use or consumption by one person, or a specific group of persons, inhibits use or consumption by one or more other persons'.[19] The Law Commission explicitly endorses this as one hallmark of a thing which attracts a property right for two reasons: first, as a rivalrous thing's capacity for use is not unlimited, competition arises as a natural consequence; second, the fact that an item is rivalrous would render the thing subject to control access, because use of the item inherently excludes another from being able to use it.[20] These two reasons reflect the core nature of the law of property, which has a primary social and economic

[15] *National Provincial Bank v Ainsworth* [1965] AC 1175 (HL) 1247–48 (Lord Wilberforce).
[16] Tatiana Cutts, 'Crypto-Property? Response to Public Consultation by the UK Jurisdiction Taskforce of the LawTech Delivery Panel' (2019) LSE Law Policy Briefing Papers 36/2019, 4 <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3406736> accessed 19 March 2023.
[17] Law Commission (n 1) para 2.39.
[18] Johan David Michels and Christopher Millard, 'The New Things: Property Rights in Digital Files?' (2022) 81 Cambridge Law Journal 323.
[19] Tatiana Cutts, 'Possessable Digital Assets: Response to the Electronic Trade Documents Law Commission Consultation Paper No. 254 and Call for Evidence on Digital Assets 2021' (2021) LSE Law Policy Briefing Paper No 47, 1 <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3895404> accessed 19 March 2023.
[20] Law Commission (n 1) para 2.68.

function to protect a person's ability to use a rivalrous thing by conferring on them property rights so as to enable them to retain access to said property. It might be noted that this criterion works in complement with excludability, the next criterion to be discussed.

C. Excludability

Another core trait of property is the factual ability of one to permit access to a thing and exclude others from its use.[21] Gray argues, however, that looking at excludability from a factual angle might not always be the most appropriate approach, and instead the proper analysis requires a *holistic evaluation*, which imbues a legal and social aspect into the overall examination.[22] Gray gives three examples. First, physical impracticability involves control over a thing, and some things are not excludable.[23] One example might include an open-air spectacle like a horse race, such as in *Victoria Park Racing v Taylor*;[24] another might be a beam from a lighthouse.[25] Second, Gray notes that 'the plaintiff who neglects to utilize relevant legal protection has failed… to raise around the disputed resource the legal fences which were available to him'.[26] Just like how the English property law accords weight to adverse possession, one might say that the failure to exercise one's right to legal protection renders a thing not excludable once the clock runs. Third, public policy might render certain things morally inappropriate to be controlled.[27] One such example would be how the law refuses to treat severed body parts as objects of property rights.[28]

But whilst important as a criterion, excludability only paints part of the picture as to what might constitute a property right. The Law Commission argues that an additional critical indicator of property rights is the criterion of separability.[29]

D. Separability

To attract property rights, the law also requires a thing to be 'subject matter independent of a person'.[30] This is illustrated by *R v Bentham*,[31] in which the House of Lords held that an unsevered hand was not a separable legal thing which could

[21] See Michael Bridge and others, *The Law of Personal Property* (3rd edn, Sweet & Maxwell 2021) para 1-006.
[22] Kevin Gray, 'Property in Thin Air' (1991) 50 Cambridge Law Journal 251, 269ff.
[23] ibid.
[24] *Victoria Park Racing and Recreational Grounds Co Ltd v Taylor* (1937) 58 CLR 479.
[25] Gray (n 22).
[26] ibid 274.
[27] Gray (n 22).
[28] See *Yearworth v North Bristol NHS Trust* [2009] EWCA Civ 37, [2010] QB 1 [30].
[29] Law Commission (n 1) para 2.72.
[30] Edelman (n 8) 53.
[31] *R v Bentham* [2005] UKHL 18, [2005] 1 WLR 1057.

be possessed. This was because 'one cannot possess something which is not separate and distinct from oneself… [and] a person's hand or fingers are not a thing'.[32] It is precisely this concept, according to Penner, which demands that intangible things such as talents, personalities, or friendships cannot be treated as property rights.[33] Penner further emphasises that what might be interesting about property rights in this regard is that 'there can be nothing special about any given property right in relation to a thing'.[34] This seems to point at elements of immutability—that one right, when transferred to another, will remain unchanged.

E. Value

The Law Commission highlights that value should be an important indicium for identifying what things should be considered as property but should not play a large role in this exercise, and suggests, in line with the power relationship formulation, that persons are more likely to seek the legal recognition and protection of valuable things than useless things.[35] In the Law Commission's view, however, a thing does not necessarily have to be imbued with value for it to attract property rights, for three reasons.[36] First, a thing which attracts property rights might not be valuable and could even attract negative value: a written-off car is at risk of incurring scrappage costs which may exceed the scrappage value, but one would scarcely say that the property rights in relation to the car are already non-existent.[37] Second, the concept of value is subjective and is at risk of volatility: value is relative, and a highly specialised item that is of great value to one might be largely worthless to another.[38] Third, information may have value, but it is not considered an appropriate object for property rights.[39]

F. Evaluation of the five criteria

The foregoing subsections have provided an overview into how the law of property has tried to wrangle definitional ambiguity into a more material framework. Insofar as the five criteria might be ranked in terms of importance, this article argues that rivalry, excludability, and separability are important in determining what constitutes property, and are furthermore inherent in the notion of a 'bundle of rights', a doctrine which this article will further dive into in the following section. As to value and the *Ainsworth* test, that such factors instead play

[32] ibid [8].
[33] JE Penner, *The Idea of Property in Law* (Clarendon Pres 1997) 112.
[34] ibid.
[35] Law Commission (n 1) para 2.80.
[36] ibid.
[37] ibid.
[38] ibid.
[39] ibid.

more of a guiding role and may not be the best way forward in bringing the law of property in line with digital assets.

Having set out the theoretical framework underpinning the idea of property, this article will now set out the broad scope of the Law Commission's proposal that a third category of personal property be developed. It will be argued that the Law Commission's proposal, although exciting, still leaves some issues to be addressed.

G. The case for digital assets as an independent class

The Law Commission argues that English law should 'explicitly recognize a third category of personal property to allow for a nuanced and idiosyncratic approach to the legal characterization of new things'.[40] As Allen and others argue, 'an analysis of the proprietary nature of digital assets' fundamentally mandates close engagement with the 'systems' they exist in, with the 'technical framework' and the 'social networks' of human actors being merely the core of the analysis.[41]

The Law Commission sets out a three-pronged test for classifying a thing as a data object. A thing is a data object if: (a) it is composed of data represented in an electronic medium, including in the form of computer code, digital, or analogue signals; (b) it exists independently of persons and exists independently of the legal system; and (c) it is rivalrous.[42] Each criterion will be examined in turn before critique is offered.

For the first criterion, the Law Commission requires that the thing in question be comprised of data which is represented in an electronic medium. The reason for this requirement is to bifurcate such assets from things in possession, which constitute of a collection of physical particles or matter within a defined boundary of three-dimensional spaces.[43] Next, they also use this criterion to acknowledge that an important part of data objects is that they have an 'informational quality' and are represented in an electronic medium which is optimised for processing by computers, and are 'uniquely instantiated' within a particular network or system.[44] In the Law Commission's view, it is the symbiotic connection between the use of specific data and the operation of 'socio-technological networks or systems' that allow said digital assets to take on characteristics or attributes that make them function more like objects than mere records.[45] Adopting such a criterion is further in line with one of the *Ainsworth* criteria—that the thing must have some form of definable or identifiable existence.[46]

[40] ibid para 4.94.
[41] ibid para 4.71.
[42] ibid para 5.10.
[43] ibid para 5.15.
[44] ibid para 5.18.
[45] ibid.
[46] ibid para 5.19.

For the second criterion, the Law Commission requires that the thing must: (a) exist independently of persons; and (b) exist independently of the legal system. This two-pronged criterion excludes things which do not have an independent existence (such as an unsevered body part) and creatures of law, such as things in action.[47] Limb (a) serves as a bastion for separability and reaffirms Michels and Millard's statement that 'to qualify as an object of property, a thing must be distinct from any person who might hold it'.[48] At the same time, it aligns with what might be implicit in *Ainsworth*, in that the object must be definable, identifiable, stable, and capable in its nature of being factually transferred to another. Viewed thus, limb (a) also deals with how a property can be asserted. A personal right can only be asserted against someone to whom it relates, whilst property rights can be asserted against the world.

Limb (b) of the second criterion requires the thing to exist independently of the legal system. This is to exclude things in action such as debt claims, which are creatures of the law which should stay in their domain.[49] This will also prevent certain statutorily created rights, such as intellectual property rights, from wandering into the domain of data objects, hence ensuring the stability of the law of property.

For the final criterion, the Law Commission requires that the thing be rivalrous, as has been discussed earlier. This criterion acts as a filter against pure information falling into the category of data objects and is in line with the proposition that 'property is rivalrous whereas information is not'.[50] Moreover, ensuring that rivalry remains an express criterion ensures that this category of objects remain consistent with the fundamental function of property law, which is to allocate rivalrous objects between individuals.

The Law Commission then tests its criteria against six different types of assets, including digital files and digital records, email accounts, certain in-game digital assets, domain names, assets connected with various types of carbon emission schemes, and crypto tokens.[51] It then provisionally concludes that not all digital assets will fall within the third category.

Ambitious as it may seem, this article argues that the Law Commission's proposal is flawed in three areas: (a) its applicability; (b) the requirement of there being the existence of data as the first criterion of its test of what constitutes a data object; and (c) its lack of clarity as to remedies. Each of these are discussed below.

[47] ibid para 5.22.
[48] Michels and Millard (n 18) 327.
[49] Law Commission (n 1) para 5.36.
[50] ibid para 5.51.
[51] ibid para 6.2.

H. Potential issues of applicability

The problems with the applicability of the Law Commission's suggested test for data objects are clear. Out of the six types of digital assets discussed by the Law Commission, only crypto tokens appear to meet the requirements of the test. It is submitted that the test is therefore myopic insofar as other digital asset classes are concerned. Only 6.2% of consumers held cryptocurrency in the United Kingdom in 2022,[52] compared to 74% of adults having sent or received emails[53] and 46% of internet users having used cloud computing services to store information in a closely surveyed period.[54] What this means is that the Law Commission's proposal as to data objects merely canvasses a niche area (that of crypto tokens), much to the detriment of many existing—and far more prevalent—digital assets. Unequal growth within the law of property would result. A proposal which only affords property rights in relation to crypto tokens and nothing else is also not a strong policy move to champion.

I. The requirement of data in the first prong of the test

This brings us on to the second critique of the test proposed by the Law Commission. As discussed earlier, the first prong of the test of whether something is a data object is whether the thing is question is 'composed of data represented in an electronic medium, including in the form of computer code, electronic, digital, or analogue signals'. This appears to draw a distinction between physical states and what exists in the digital realm, creating the false impression that there might be objects which are only represented through 'analogue data' but lack a coded iteration. As Cutts puts it, the proper approach is that the inquiry ought instead consider 'how the characteristics of those assets are described and communicated by individuals operating within the systems that we use to deal with them, rather than the physical changes that those characteristics would cause'.[55] Indeed, Cutts points out that it is the code which represents the 'assets' we deal with in the digital realm, rather than 'the values of certain physical states that running the code may precipitate at any given moment'.[56]

[52] TripleA, 'United Kingdom: Cryptocurrency Information about the UK' (*TripleA*, 2022) <https://triple-a.io/crypto-ownership-united-kingdom-2022/#> accessed 19 March 2023.
[53] Anyi Petrosyan, 'Share of internet users who sent and received emails in the United Kingdom (UK) in 2020, by age group' (*Statista*, 29 August 2022) <https://www.statista.com/statistics/506315/sending-and-receiving-emails-in-the-united-kingdom-uk-by-age-group/> accessed 21 March 2023.
[54] Lionel Sujay Vailshery, 'Cloud computing in the United Kingdom – Statistics & Facts' (*Statista*, 3 February 2022) <https://www.statista.com/topics/3164/cloud-computing-in-the-united-kingdom-uk/#topicOverview> accessed 21 March 2023.
[55] Tatiana Cutts, 'Assets Represented by Computer Code: Response to "Digital Assets: Law Commission Consultation Paper 256"' (2022) LSE Law Policy Briefing Paper No 50, 3 <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4259516> accessed 19 March 2023.
[56] ibid.

Thus, instead of questioning whether something is 'composed of data represented in an electronic medium', it might be more apposite to ask whether it is 'composed of data represented in computer code'.[57] Such a test properly reflects what the end-user interacts with, instead of abstract binary physical states which might exist on electronic mediums. As the world embraces virtual reality and machine learning begins to grow exponentially, an artificial focus on the 'electronic' or 'digital' implications of holding data might not be the best way forward. A tweak of the first part of the test is therefore necessary.

J. Remedies

The issue of remedies is one of the most important in any legal framework, and there are currently three main issues which render the Law Commission's proposal potentially unworkable.

First, in creating a third category in property, the Law Commission points out that various existing legal frameworks could be applied to data objects, such as breach of contract, following and tracing, restitutionary claims, and the like.[58] The Law Commission discusses how these remedies might apply to crypto tokens, instancing an example of proprietary restitution and the possible extension of the tort of conversion.[59] Yet the Law Commission omits to state how these remedies might apply to other digital assets. Insofar as other digital assets such as domain names and digital files are concerned, it remains unclear as to what legal remedies ought to be accorded in the event of a dispute relating to said assets. This is an issue which has cropped up across various jurisdictions. For example, in both England and British Columbia, disputes have emerged over over ownership of domain names.[60] Likewise, disputes have raged on both sides of the Atlantic with regard to ownership over digital files and access to emails.[61] How remedies such as a proprietary restitutionary claim might apply to other types of digital assets other than crypto tokens remain to be seen.

Second, the Law Commission argues that tracing (rather than following) provides the correct analysis of the process which ought to apply to locate and identify the claimant's property when said crypto tokens are transferred.[62] It has arguably erred in this respect. Cutts correctly argues that tracing and following are distinct concepts: tracing, as she puts it, is about 'characterising transactions by which [one identifies] substitute assets', whilst following is about 'pursuing assets

[57] ibid.
[58] Law Commission (n 1) para 19.88.
[59] ibid paras 19.73–19.76.
[60] *Hanger Holdings v Perlake Corporation SA* [2021] EWHC 81 (Ch), [2021] Bus LR 544; *Canivate Growing Systems Ltd v Brazier*, 2020 BCSC 232.
[61] As to domain names, see *Your Response Ltd v Datateam Business Media Ltd* [2014] EWCA Civ 281, [2015] QB 41; *Thyroff v Nationwide Mut Ins Co* 864 NE2d 1272, 1273 (NY 2007). As to access to digital files, *Fairstar Heavy Transport v Adkins* [2013] EWCA Civ 886; *Ajemian v Yahoo!, Inc* 84 NE3d 766 (Mass 2017).
[62] Law Commission (n 1) paras 19.47–19.52.

from one location to another'.[63] Two distinct types of cases seem to be ubiquitous in the case law. The first type involves an 'asset substitution in one set of hands', and the second type involves a 'bank transfer from one account to another'.[64] As Cutts argues, there is an underlying distinction between these two types of cases. In the context of bank transfers, said claims are traditionally 'reified', wherein the courts treat these cases as though they involve a transfer of an asset independent of the underlying account.[65] It is here where Cutts's 'dummy asset tracing' theory sheds light on the weakness of the Law Commission's proposal. Cutts argues the weakness of the bank transfer cases lies in that the courts are scarcely dealing with anything related to substitution, for courts are merely 'following a fictional cash asset from one location to another'.[66] The justification propounded by courts for engaging in this practice of 'dummy asset tracing' is usually that there would be liability if the facts had involved some dealings in physical monies.[67] Respectfully, this proposition is unjustifiable at both the individual and institutional level. For the former, as opposed to physical objects and coded entities (such as data assets), bank funds do not have any strict or visible parameters, and there is nothing much that a payee may do to discover a prior claim.[68] Insofar as there are tenable arguments for reversing a defective transaction, such arguments scarcely extend to recovery of funds against one who might not be privy to a transaction.[69] For the latter, as Cutts observes, the 'irrevocability of payment instructions' already provides any confidence that is needed for the free circulation of money.[70] These cases provide weak judicial grounding for the application of tracing in relation to crypto tokens, as the doctrine of 'dummy asset tracing' has led to an unyielding complexity of cases which involve tracing through multiple accounts, which might instead be better dealt with by 'standard principles of characterization'.[71]

Indeed, setting aside the dummy transaction doctrine does not render the current law otiose as it relates to crypto tokens. This is because transfer of crypto tokens can be subject to characterization through the doctrine of following. Given that crypto tokens operate in the domain of legal assets, no issues arising from a change in physical form (and thereby to changes in ownership because of specification) will arise. Instead, the main question would be the extent of the protection which we wish to accord to the original owners of crypto assets. The defence of innocent purchase can play a larger role in this picture, to coordinate the evolved ecosystem between crypto tokens, following, and equitable remedies.

Third, it is contended that the Law Commission has taken an unnecessarily narrow view to the extension of the tort of conversion. The Law Commission

[63] Cutts (n 55) 7.
[64] ibid.
[65] ibid.
[66] ibid.
[67] ibid 8.
[68] ibid.
[69] ibid.
[70] ibid.
[71] ibid 7.

argues that 'there is international precedent for extending the tort of conversion to objects of property rights that would fall within [the third category of personal property]' but concedes that such an extension would not extend to certain data objects such as digital files.[72] This concession should be reconsidered. This is because other jurisdictions have decided to extend the tort of conversion to domain names and digital files.[73] Apart from crypto tokens, therefore, it is also important for the tort of conversion to be extended to other digital assets, and this is something that the Law Commission should have considered.

K. An overall evaluation

Section II of this paper has evaluated the Law Commission's proposals as to property rights in digital assets, after exploring the basic concepts of the law of property. The current position adopted by the Law Commission remains myopic as to how it applies digital assets other than crypto tokens. Indeed, legal uncertainty continues to loom large as it relates to such assets, especially in the context of cloud storage and other intangibles. Further, the Law Commission's criteria for 'data objects'—particularly its definition of data—and associated legal remedies could be further reworked. Building on the Law Commission's proposal, Section III of this article then seeks to explore how common law jurisdictions have incorporated digital assets within the law of property, particularly through evaluating into the policy set-up and the juridical technological discourse that has occurred to date.

III. The Juridical Technological Discourse as it Relates to Cryptocurrencies and NFTs

Writing in 1991, Gray believed that 'before long [I] would have sold you a piece of thin air and you will have called it property'.[74] In the present day, Gray's surprisingly prescient statement has somehow morphed into reality. Today, a piece of 'thin air', such as an NFT, could very well be worth more than a physical copy of the same design. The paradigm ushered in by the recent NFT craze has therefore forced property lawyers and courts (and indeed the English Law Commission) to re-examine the substantive foundations of property law. Section III of the article will examine case law from different common law jurisdictions and argue that the approach taken by the Singapore High Court in *Janesh s/o Rajkumar v Unknown Person* is commendable and offers to usher in stable guidance for the English Law Commission, particularly as it develops its test for 'data objects'.[75]

[72] Law Commission (n 1) para 19.113.
[73] See, for example, *Kremen v Cohen* 337 F3d 1024 (9th Cir, 2003) and *Henderson v Walker* [2019] NZHC 2184.
[74] Gray (n 22) 252.
[75] *Janesh s/o Rajkumar v Unknown Person* [2022] SGHC 264.

A. The policy set-up

The fundamental policy set-up within which the juridical technological discourse takes place stems from how crypto assets offer a potential counter against what Zuboff terms as the 'surveillance economy'.[76] Financial transactions form a critical cog within the privacy ecosystem, for they reveal potentially huge amounts of information regarding the volume and transactions of purchases, location histories, and even social networks. Therefore, crypto tokens have been seen as a supposedly 'privacy enhancing' mechanism which serves to improve the security and reliability of transactions. Indeed, the exchange of assets without the need for a centralised financial institution, underpinned by a distributed ledger system which is a product of autonomous computers, does present huge potential for the modern economy, and has led regulators to believe that the 'next wave' of technological evolution is in crypto tokens.

What crypto tokens serve to do would be to avoid the current problems engendered by the banking system. When a transaction is made using paper currency, all that a receiver has to do is to check that the currency is not counterfeit. In the case of digital transactions, the authentication is done by an intermediary like a bank, as most electronic transfers are done by one bank to another. Crypto tokens essentially attempt to remove this intermediary altogether by separating all trust institutions and creating a private ecosystem which is self-regulated.[77] The mechanisms adopted by crypto tokens—namely, distributed ledger technology, the authentication of transactions, and the ability to send and receive payments directly—reflect the emergence of a banking system of the future. This is what has led to the surge in interest with regard to such assets, and its increased adoption has led to increased debate across society as to how it might be regulated. This debate has trickled into the juridical discourse, to which we now turn.

B. The US

State law in the US has remained unclear as to whether cryptocurrencies *should* be treated as property. At the District Court level, *Currier v PDL Recovery Group, LLC* involved a case where a creditor had filed a request in a bid to liquidate BTC and ETH tokens held by the defendant on a crypto exchange.[78] The court ruled that '[its] ability to order satisfaction of a judgment with a defendant's personal property that is in possession of a third party is limited'. Simply put, the Court considered that the crypto tokens held by the defendant were intangible

[76] Shoshana Zuboff, 'Big Other: Surveillance Capitalism and the Prospects of an Information Civilisation' (2015) 30 Journal of Information Technology 75.
[77] T Rabi Sankar, 'Cryptocurrencies—An Assessment' (Bank for International Settlements 14 February 2022) <https://www.bis.org/review/r220217d.pdf> accessed 19 March 2023.
[78] *Currier v PDL Recovery Grp LLC*, No 14-12179, 2018 WL 4057394 (ED Mich Aug 27, 2018).

personal property and therefore, a liquidation order was denied to the plaintiff. Likewise, *Rasmussen v Smith* came to the same conclusion.[79] *Rasmussen* involved a court-appointed receiver suing various defendants to recover various crypto tokens. The Texas State Court granted summary judgment to the plaintiff and agreed that the original owner had property rights in the crypto tokens. Taking these cases together, juridical discourse at both the District and State level in the US seem to suggest a bright-line rule that cryptocurrencies can be considered as property.

The aforementioned cases, however, chafe uneasily against other District Court rulings, such as *Temurian v Piccolo*.[80] There, the Florida District Court denied the plaintiff's claim in conversion in relation to crypto tokens. Conversion in Florida law is defined as 'the wrongful exercise of dominion or control over property to the detriment of the rights of one entitled to possession'.[81] An action for conversion of money consists of three elements: (a) specific and identifiable money; (b) a deprivation of money belonging to another; and (c) an unauthorised act, which deprives the other of their money. The court considered that the Eleventh Circuit (Federal Law) had yet to decide whether cryptocurrencies were considered 'money' for the purposes of conversion. Although there had been cases where courts recognised cryptocurrencies as considered 'money' under the ambit of several federal money laundering statutes,[82] the court considered that even if this was the case, for the purposes of conversion 'money [must be] in a specifically identifiable fund such as an escrow account, a bag of gold coins, and the like'.[83]

Other courts, meanwhile, have developed rules relating to liability. In *Day v Boyer*, a Californian State Court awarded damages to a plaintiff who had purchased various crypto tokens but did not receive them.[84] Likewise, in *Smoak v Bitcoin Market*, a temporary denial of access to a plaintiff's wallet at a crypto exchange led to the grant of a default verdict by the Oklahoman State Court with damages calculated by reference to Bitcoin's price at the point in time when access was blocked.[85] Finally, in *Rensel v Centra Tech*, it remains notable that the court decided not to adopt a proprietary analysis, and instead awarded the plaintiffs damages without considering the possibility of ordering the return of the specific tokens transferred to the plaintiffs.[86]

Whilst there has yet to be any bright-line judicial statement at the Federal Court level affirming that cryptocurrencies can be considered a form of property, various guidance by top-level US regulators suggest that the US may move in such

[79] *Rasmussen v Smith*, No 3:18-CV-01034-M, 2020 WL 109863 (ND Tex Jan 8, 2020).
[80] *Temurian v Piccolo*, No 18-CV-62737, 2019 WL 1763022 (SD Fla Apr 22, 2019).
[81] *United States v Bailey*, No 6:01-CV-875-Orl-22KRS, 288 F Supp 2d 1261, 1269 (MD Fla 2003).
[82] *United States v Faiella*, 39 F Supp 3d 544, 545 (SDNY 2014); *SEC v Shavers*, No 4:13-CV-416, 2013 WL 4028182 (EDTex Aug 6, 2013).
[83] *Talisman Capital Alt Invs Fund, Ltd v Mouttet*, No 12-01842-LMI, 493 BR 640, 662 (Bankr SD Fla 2013).
[84] *Day v Boyer*, No 19-CV-01669, 2020 US Dist LEXIS 9959 (CD Cal Jan 21, 2020).
[85] *Smoak v Bitcoin Market, LLC*, No CIV-18-1096-PRW (WD Okla Jul 24, 2019).
[86] *Rensel v Centra Tech, Inc*, No 17-24500-CIV-KING/SIMONTON, 2018 US Dist LEXIS 106642 (SD Fla June 25, 2018).

a direction. Moving towards such a position is attractive from a public policy point of view, particularly given that the US is home to the largest number of crypto investors, exchanges, trading platforms, crypto mining firms, and investment funds.[87]

In this regard, the Securities and Exchange Commission (SEC) often views many crypto assets as securities. Indeed, the Commodity Futures Trading Commission (CTFC) calls Bitcoin a commodity[88], and the Treasury calls it a currency.[89] Going further, the Internal Revenue Service (IRS) defines cryptocurrencies as a 'digital representation of value that functions as a medium of exchange, a unit of account, and/or a store of value', requires investors to disclose their yearly cryptocurrency activities on tax returns, and has accordingly issued tax guidance.[90] The IRS also treats virtual currency and property for the purposes of US Federal Tax and applies the general rules for property transactions.[91] Such an approach might very well extend to other areas in US jurisprudence and might serve to inform the development of federal law in this context.

Absent express judicial guidance, the historical understandings of the foundations of property law might usefully be examined so to understand the current juridical discourse. There are two main views on the right to property in the US. Smith and Merrill illustrate the constant duel between traditionalists and supporters of the bundle of rights view; the former believe there is a core, inherent meaning in the concept of property, whilst the latter argue that a property owner only has a bundle of permissive uses over the property.[92] Traditionalists largely argue that three rights—the right to exclusion, the right to use, and the right to transfer—define property. In contrast, proponents of the bundle of rights view tend to argue that property is a bundle of rights defined by law and public policy, but that what remains in the bundle of rights is a matter of policy and the content of the right is inconsequential.[93]

This article argues that the bundle of rights view should find favour, particularly in the context of digital assets. Pioneered by Hohfeld, the bundle of rights view is underpinned by the theory that property does not consist of things, but

[87] Susannah Hammond and Todd Ehret, 'Cryptocurrency Regulations by Country' (Thomson Reuters 2022) 5 <https://www.thomsonreuters.com/en-us/posts/wp-content/uploads/sites/20/2022/04/Cryptos-Report-Compendium-2022.pdf> accessed 19 March 2023.

[88] US Commodity Futures Trading Commission, 'Bitcoin Basics' (US Commodity Futures Trading Commission 2018) 1 <https://www.cftc.gov/sites/default/files/2019-12/oceo_bitcoinbasics0218.pdf> accessed 21 March 2023.

[89] US Department of the Treasury, 'Frequently Asked Questions' (*US Department of the Treasury*, 2023) <https://home.treasury.gov/policy-issues/financial-sanctions/faqs/topic/1626> accessed 21 March 2023.

[90] Internal Revenue Service, 'IRS Virtual Currency Guidance: Virtual Currency is Treated as Property for U.S. Federal Tax Purposes; General Rules for Property Transactions Apply' (*Internal Revenue Service*, 25 March 2014) <https://www.irs.gov/newsroom/irs-virtual-currency-guidance> accessed 19 March 2023.

[91] ibid.

[92] Thomas W Merrill and Henry E Smith, *The Oxford Introductions to US Law: Property* (Oxford University Press 2010).

[93] Richard A Epstein, *Takings: Private Property and the Power of Eminent Domain* (Harvard University Press 1985).

instead fundamental legal relations between people. Ownership, according to Hohfeld, is not binary, characterised by the simple and non-social relationship between a person and a thing, but instead is better viewed as a 'complex set of social relations in which individuals are interdependent'.[94] Whether and how courts should make value choices about what property law ought to prefer should depend on the 'policies society [has] decided to promote'.[95] This view seems to chime in accord with the Law Commission's position as explored in Section II, which holds that property as 'not a thing at all', but instead a 'socially approved power relationship'.[96] This view is particularly important with the advent of digital assets, as the policy set-up has shown that such objects have the potential to change the way that modern society fundamentally operates. The three-pronged test for data objects developed by the Law Commission (particularly when dealing with the factors of rivalry, excludability, and the like) are reflective of what deserves protection in the modern world, and crypto tokens do fulfil this test and deserve protection.

Certainly, in the US, it is likely that the bundle of rights view will find favour such that cryptocurrencies and NFTs are recognised as property, particularly given the direction that various top-level regulators have been moving towards in reining such digital assets into the definition of 'property'. Whilst there has yet to be any US case ruling expressly on whether NFTs can be considered property, it is likely that normative principles will continue to underpin the analysis.

C. ENGLAND

Across the Atlantic, the judicial position as it relates to cryptocurrencies and NFTs is far less ambiguous. Indeed, the English High Court in *AA v Persons Unknown* has mostly swept any uncertainty formerly brewing in English law as it relates to cryptocurrencies.[97]

AA v Persons Unknown involved a dispute tackling the question of whether Bitcoin can constitute 'property' which was capable of being a subject of a proprietary injunction. In short, it involved a defendant who had infiltrated the security systems of an insured customer using malware, which caused the forced encryption of all the computer systems. The defendants then offered the plaintiffs a decryption tool, upon the transfer of over USD $1.2 million worth of Bitcoin. This was essentially an act of blackmail, although the incident response company instructed by the plaintiff did eventually manage to negotiate the ransom down to USD $950,000 (109.25 Bitcoins), a sum which was paid.

The argument before the High Court ultimately centred around the issue of whether a proprietary injunction should be granted over the transferred

[94] Wesley Newcomb Hohfeld, 'Some Fundamental Legal Conceptions as Applied in Judicial Reasoning' (1913) 23 Yale Law Journal 16, 19. See also Denise R Johnson, 'Reflections on the Bundle of Rights' (2007) 32 Vermont Law Review 247, 251.
[95] Johnson (n 94) 251–252.
[96] Law Commission (n 1) para 2.10.
[97] *AA v Persons Unknown* [2019] EWHC 3556 (Comm), [2020] 4 WLR 35.

Bitcoins. To succeed in obtaining a proprietary injunction, the plaintiffs had to first show that the cryptocurrencies were property. Prior to *AA*, the judicial position in England was much like the US, in that a clear bright-line rule had yet to take shape. In September 2018, the High Court in *Elena Vorotyntseva v Money-4 Service Ltd* granted a freezing order over BTC and ETH.[98] The court held that there was no suggestion that cryptocurrency 'cannot be a form of property or that a party amenable to the court's jurisdiction cannot be enjoined from dealing with or disposing of it'.[99] Although this case was favourable to the argument that cryptocurrencies are indeed property, scarce reasoning was given. Likewise, in the unreported case of *Robertson v Persons Unknown*, cryptocurrencies were once again viewed by the court as property.[100] This case involved an application by the claimant for an asset preservation order (APO) to secure the 80 BTC and a Bankers Trust order to reveal the identity of the wallet holder. The Court acknowledged that various difficulties came before it, including, but not limited to, the fact that Bitcoin was neither a property that a party can take physical possession of, nor did it create a property right which could be obtained or enforced through legal action. However, the court skirted around this, invoking the Singapore Court of Appeal case of *Quoine v B2C2* to support the view that cryptocurrencies constituted property.[101]

Scarce reasoning plagued English law until *AA v Persons Unknown*. In *AA*, Bryan J recognised that the difficulty in treating cryptocurrencies as property was that they fell neither into the categories of choses in action nor choses in possession.[102] Cryptocurrencies, according to Bryan J, did not constitute the former because they did not embody any right that was capable of being enforced by action, and did not constitute the latter because they were neither tangibles nor can they be possessed.[103] Therefore, they sat uneasily against Fry LJ's seminal statement in *Colonial Bank v Whinney*, that 'all personal things are either in possession or action. The law knows no *tertium quid* between the two'.[104]

To rebut Fry LJ's position, Bryan J relied on the UK Jurisdiction Taskforce's (UKJT) Legal Statement on Cryptoassets and Smart Contracts. [105] According to Bryan J, it was 'fallacious to proceed on the basis that the English law of property recognizes no forms of property other than choses in possession and choses in action'.[106] Indeed, insofar as *Colonial Bank* stood as a bastion against the proposition that cryptocurrencies should be recognised as property, the UKJT

[98] *Elena Vorotyntseva v Money-4 Ltd (t/a Nebeus.com)* [2018] EWHC 2596 (Ch).
[99] ibid [13].
[100] *Robertson v Persons Unknown* (Com Ct, 15 July 2019).
[101] *Quoine Pte Ltd v B2C2 Ltd* [2020] SGCA(I) 02.
[102] *AA* (n 97) [55].
[103] ibid.
[104] *Colonial Bank v Whinney* (1885) 30 Ch D 261 (CA) 285.
[105] UK Jurisdiction Taskforce, 'Legal Statement on Cryptoassets and Smart Contracts' (The LawTech Delivery Panel November 2019) <https://resources.lawtechuk.io/files/4.%20Cryptoasset%20and%20Smart%20Contract%20Statement.pdf> accessed 19 March 2023.
[106] *AA* (n 97) [58].

reasoned that 'it is not clear…whether Fry LJ intended this [narrow dualistic] view',[107] and that the Court of Appeal in *Colonial Bank* did not 'explicitly address the issue of exhaustive classification between things in action and things in possession and said nothing about the definition of property'.[108] Hence, Bryan J was able to conclude that *Colonial Bank* was not to be treated as 'limiting the scope of what kinds of things can be property in the law'.[109]

Another hurdle, however, was *Your Response v Datateam Business Media*.[110] There, Moore-Bick LJ said that *Colonial Bank* made it 'very difficult to accept that the common law recognizes the existence of intangible property other than [things] in action', but even if it did, the decision in *OGB Ltd v Allan*[111] 'prevents [the court] from holding that property of that kind is susceptible of possession so that wrongful interference can constitute the tort of conversion'.[112] Although Moore-Bick LJ considered that there 'was a powerful case for reconsidering the dichotomy between [things] in possession and [things] in action and recognizing a third category of intangible property', the court held that it was not allowed to do so because of *OGB*.[113]

Notwithstanding, the UKJT noted that *Your Response* did not stand for the proposition that intangible things other than things in action could never be property at all. Indeed, the only proposition it stood for was that they could not be the subject of certain remedies. One must distinguish between a database containing purely information, which was the subject matter in *Your Response*, and intangible assets with special characteristics, as may be the case in cryptoassets.[114] Likewise, in *Swift v Dairywise Farms Ltd*, the court held that a milk quota could be the subject of a trust;[115] whilst in *Armstrong v Winnington*, the court held that EU carbon emissions could be the subject of a tracing claim as a form of 'other intangible property', even though it was neither a thing in possession nor a thing in action.[116] Indeed, other English statutes define property in terms which assume that intangible property is not limited to things in actions as well.[117]

In *AA*, Bryan J therefore affirmed the view of the UKJT and the that of Singapore Court of Appeal in *Quoine v B2C2*, and concluded that cryptocurrencies properly constituted property under English law as they satisfied Lord Wilberforce's four criteria in *Ainsworth*, set out in Section II.A above.[118] It should nevertheless be added, however, that *AA v Persons Unknown* was decided on an

[107] UK Jurisdiction Taskforce (n 105) para 74.
[108] ibid para 76.
[109] ibid para 77.
[110] *Your Response* (n 61).
[111] *OGB Ltd v Allan* [2007] UKHL 21, [2008] 1 AC 1.
[112] *Your Response* (n 61) [26].
[113] ibid [27].
[114] UK Jurisdiction Taskforce (n 105) para 81.
[115] *Swift v Dairywise Farms Ltd* [2001] EWCA Civ 145, [2001] 1 BCLC 672.
[116] *Armstrong DLW GmbH v Winnington Networks Ltd* [2012] EWHC 10 (Ch), [2013] Ch 156.
[117] UK Jurisdiction Taskforce (n 105) para 83.
[118] *AA* (n 97) [59].

interim application, and hence the claimant only needed to reach the threshold that there was a serious issue to be tried.

Academics such as Bridge and Gullifer are likely to approve of the result of *AA v Persons Unknown*. Indeed, they acknowledge that whilst one view is that there exist only two categories of personal property, this being Fry LJ's two categories as set out in *Colonial Bank*, there also exists another category of 'intangible property'.[119] Indeed, they highlight that recent developments in relation to intangible property generally, and more specifically crypto assets, have reignited the debate and support such a view.[120]

In *Lavinia Deborah Osborne v Persons Unknown*, the reasoning in *AA v Persons Unknown* was further extended to NFTs, albeit not in the level of detail as was addressed by Bryan J.[121] The claimant asserted that two digital artworks from a particular NFT collection, which she had purchased through an NFT marketplace, had been stolen from her online digital wallet. The court held that there was a realistically arguable case that NFTs could be treated as property under English law.[122] The court's position was that there was no other reason to treat NFTs in any other different way and assumed as a matter of English law that they were to be treated as property.[123]

Whilst cases remain scarce, it is submitted that *AA v Persons Unknown* sets a solid foundation for the juridical discourse relating to cryptocurrencies and NFTs. Nevertheless, a comparative study would offer valuable insights which may serve to pave the way forward for cross-fertilisation between common law jurisdictions.

D. New Zealand

The New Zealand case of *Ruscoe v Cryptopia* offers valuable judicial insights relating to cryptocurrencies and NFTs.[124] The nub of the argument in this case centred on whether cryptocurrencies were property for the purposes of section 2 of New Zealand Companies Act 1993.[125] If they were, they would then be held on trust for account holders following the liquidation of a company. Section 2 defined 'property' as 'property of every kind whether tangible or intangible, real or personal, corporal or incorporeal, and includes rights, interests, and claims of every kind in relation to property however they arise'.[126] It was accepted by Gendall J that 'property' in the context of the Act was a 'wide' concept which included 'money', even though that was not expressly included in within the section itself.[127]

[119] Bridge and others (n 21) para 1-107.
[120] ibid.
[121] *Lavinia Deborah Osborne v Persons Unknown* [2022] EWHC 1021 (Comm).
[122] ibid [13].
[123] ibid [15].
[124] *Ruscoe v Cryptopia Ltd* [2020] NZHC 728, [2020] 2 NZLR 809.
[125] ibid [70].
[126] Companies Act 1993 (NZ), s 2.
[127] *Ruscoe* (n 124) [70]–[75].

In *Ruscoe*, the court was firm in holding that cryptocurrencies were property for the purposes of New Zealand law. Indeed, like Bryan J in *AA v Persons Unknown*, the court approved of the contents of the UKJT statement.[128] According to the court, there were heavy public policy reasons against ruling otherwise, because this would have unsatisfactory implications for 'New Zealand's law, including insolvency law, succession law, law of restitution, and commercial law'.[129]

In setting the stage up for the proposition that cryptocurrencies constituted a form of property, the court cited the Singapore case of *Quoine v B2C2*.[130] This case involved a claim for breach of trust, which could only succeed if the bitcoins in question were an asset that could form the subject matter of the trust. On the 'property' question, the Singapore Court of Appeal however declined to make an affirmative decision as to whether cryptocurrencies could constitute a form of property. Instead, Menon CJ commented that 'there may be much to commend the view that cryptocurrencies should be capable of assimilation in the general concepts of property… [however] there are questions as to the type of property that is involved'.[131] Despite the tentative nature of the proposition, it clearly gave judicial support for the argument that cryptocurrencies can constitute a form of property, and *Ruscoe* certainly latched on and further developed on this proposition.

The court in *Ruscoe*, in exploring the boundaries of the legal concept of "property", further discussed two New Zealand cases. The first of these was *Dixon v R*. There, the New Zealand Supreme Court held that a digital copy of a CCTV footage was 'property' in the context of section 2 of the Crimes Act 1961 seemed to endorse the view that computer data would meet the general definitions of property.[132] Indeed, the reason why the digital footage was not merely "information" in *Dixon* was because it could be identified, had a value, was capable of being transferred, and had a physical presence, albeit one that could not be detected by the means of unaided sensors.[133] The second was *Henderson v Walker*.[134] There, Thomas J held that in principle, a common law action in conversion was available with respect to certain conduct which had occurred in relation to computer data. The digital files were both excludable and exhaustible, and were therefore capable of cognitive and manual control—both essential requirements for the tort of conversion.[135] As to excludability, this was because digital files had a material presence, which could physically alter the medium on which they are held. This physical presence, according to the court, allowed others to be excluded from the digital asset, either by physical control of the medium or by using password

[128] ibid [64]–[65].
[129] ibid [66].
[130] ibid [77]–[83].
[131] *Quoine* (n 101) [144].
[132] *Dixon v R* [2015] NZSC 147, [2016] 1 NZLR 678 [51].
[133] *Ruscoe* (n 124) [95]–[96].
[134] *Henderson v Walker* [2019] NZHC 2184.
[135] ibid [264], [266].

protection.[136] As to exhaustibility, the court held that digital files could be deleted or modified to render them useless or inaccessible.[137] In *Ruscoe*, Gendall J considered that the reasoning of Thomas J could reasonably be extended to wrongful interferences with cryptocurrencies or digital assets, as any person who had gained unauthorised access to the private key attached to cryptocurrencies and used them 'would permanently deprive the proper possessor of the [cryptocurrencies] of that property and its value".[138] On the basis of these two cases, Gendall J was therefore able to reach the view that the proposition that information is not 'property' did not apply to where digital assets were concerned. [139]

What, then, might be the appropriate threshold test to determine whether something constituted property within New Zealand? Gendall J ruled that the *Ainsworth* test ought to apply and applied Lord Wilberforce's four criteria to the cryptocurrencies in issue.

First, the asset had to be definable. In other words, the asset must be capable of being isolated from other assets, whether of the same or of other types and identified. Gendall J held that this was satisfied for cryptocurrencies. This was because 'computer-readable strings of characters recorded on networks of computers established for the purpose of recording those strings… [were] sufficiently distinct to be allocated to a particular accountholder on the network'.[140] The cryptocurrencies involved in the present case contained a public key, which was responsible for allocating each string to a unique user. The working of the system, according to Gendall J, was that 'the distribution of the data across a large network of computers, when combined with cryptography that prevents individual networks from altering historical data over the network, assists in giving that data stability'.[141] Thus, viewed as a whole, cryptocurrencies were certainly definable. Furthermore, the public key allocated to a cryptocurrency can be viewed as more identifiable than some asserted rights. There is therefore a compelling reason to accept that cryptocurrencies would be able to fulfil the first limb of the *Ainsworth* test.

The second requirement of *Ainsworth* is that the asset needs to be identifiable by third parties. According to Gendall J, 'the unique strings of data recording the creation and dealings with cryptocurrency are always allocated via the public key to a particular accountholder connected to the system', and in the context of cryptocurrencies:

> [t]he degree of control necessary is achieved… by the computer software allocating to each public key a second set of data made available only to the holder of the account (the private key) and requiring the

[136] ibid [265].
[137] ibid [266].
[138] *Ruscoe* (n 124) [93].
[139] ibid [98].
[140] ibid [105].
[141] ibid.

> combination of the two sets of data in order to record a transfer of the cryptocurrency attached from one public key to another.[142]

These features therefore prohibit involuntary transfers and the ability to transfer the cryptocurrency data twice.[143] In this regard, the second limb of *Ainsworth* is likely satisfied.

The third requirement of *Ainsworth* is that the right must be capable of assumption by third parties. Gendall J viewed that cryptocurrency met this requirement, and the fact that they were the subject of active trading markets furthered such an argument.[144]

The fourth requirement of *Ainsworth* is that the thing in question must have some degree of permanence or stability. Gendall J opined that the 'blockchain methodology which cryptocurrency systems deploy… greatly assist in giving stability to cryptocoins', and that '[t]he entire life history of a cryptocoin is available in the public recordkeeping of the blockchain'.[145] Indeed, a particular cryptocoin is in 'existence and stable until it is spent through the use of the private key, which may never happen [as] standard cryptocurrency systems do not provide for the arbitrary cancellation of coins'.[146] Viewed thus, cryptocurrencies are likely to fulfil the fourth limb of *Ainsworth* as well.

It remains to add that the public policy argument that 'some types of cryptocurrencies are used by criminals for the transmission of funds across borders and as a means of laundering the proceeds of past criminal activity' was viewed by Gendall J as an unpersuasive argument against recognising cryptocurrencies as property, as such issues were not unique to cryptocurrencies, and the increasing use by the traditional banking sector of cryptocurrencies was indicative of a need to recognise such assets as property to spur commercial development.[147]

E. AUSTRALIA

The position in Australia supports the proposition that digital assets are indeed property. In *Hauge v Cordiner (No 2)*, the New South Wales District Court approved the claimant's cryptocurrency investment reserves (which were in Bitcoin) as security for costs.[148] Whilst this was opposed by the defendant because cryptocurrencies were a highly unstable form of investment,[149] Gibson DCJ contended that cryptocurrencies, although 'volatile', were 'a recognized form of

[142] ibid [109]–[112].
[143] ibid [113].
[144] ibid [114]–[116].
[145] ibid [118].
[146] ibid.
[147] ibid [129]–[130].
[148] *Hauge v Cordiner (No 2)* [2020] NSWDC 23.
[149] ibid [30].

investment'.[150] He cited *Noicos v Dawson* in support of this argument.[151] In this case, White J noted that although the applicants for the freezing orders were cryptocurrency investment dealers, they were nonetheless considered to be able to offer an undertaking as to damages in relation to injunctive relief.

In all, therefore, common law jurisdictions are beginning to recognise cryptocurrencies as a credible source of value, and there has been a subtle push towards the position that cryptocurrencies—and other digital assets—are indeed property.

F. Singapore

Having conducted an overview of the law in most common law jurisdictions, we are left with one overhanging question. *Prima facie*, it appears that most common law jurisdictions are open towards recognizing cryptocurrencies and NFTs as a form of property, but the route that courts have taken towards that conclusion diverge. Insofar as the *Ainsworth* test (which has been adopted by the English and New Zealand courts) provides a solution, its appropriateness remains debatable. This article argues that the recent Singapore High Court case of *Janesh s/o Rajukumar v Unknown Person* in the Singapore High Court provides a timely development in this flourishing area of the law.[152] The ruling of this case is significant, as it marks the first instance in Asia where a court has explicitly recognised an NFT as a form of legal property and digital assets with proprietary rights attached to them.

In *Janesh*, the claimant was the proud owner of an NFT known as the Bored Ape Yacht Club ID #2162 (hereinafter 'Bored Ape NFT'). The claimant acquired the NFT when he purchased it for 15.99 ETH on Opensea, an online NFT marketplace, on 6 August 2021. He was a regular user of NFTfi, a community platform functioning as an NFT-collateralised cryptocurrency lending marketplace. One NFT he often used as collateral was said Bored Ape NFT due to its rarity and value.[153] Whenever he used the NFT as collateral, the claimant was careful to specify that: (a) the Bored Ape NFT would be transferred to NFTfi's escrow account until full repayment of the loan was effected; (b) in the event that the claimant was unable to make full repayment of the loan on time, he would inform the lender who should provide reasonable extensions of time for repayment; (c) at no point should the lender use the 'foreclose' option of NFTfi's Smart Program on the Bored Ape NFT without first granting the claimant reasonable opportunities to make full repayment of the loan and retrieve the Bored Ape NFT from the escrow account; and (d) at no point would the lender obtain ownership, nor any right to sell or dispose of the Bored Ape NFT.[154] The lender could only, at best, hold on to the Bored Ape NFT, pending repayment of the loan.[155]

[150] ibid [31].
[151] *Noicos v Dawson* [2019] FCA 2197.
[152] *Janesh* (n 75).
[153] ibid [11].
[154] ibid.
[155] ibid.

Things went well for the claimant, until he reached out to the defendant sometime around January 2022 for a loan for 45 ETH. This was for a period of 90 days, with interest payable at 33% per annum. This loan was eventually repaid.[156] In March 2022, the defendant offered the claimant another loan, for another 150,000 DAI (an alternate form of cryptocurrency). This was for 30 days with interest payable at 45% per annum.[157] In April 2022, the claimant told the defendant that he needed some more time to repay the loan, which the defendant agreed. Thereafter, the claimant informed the defendant two days later that he had reached out to another user to repay the outstanding amount, and the defendant thereafter agreed to the new refinancing loan arrangement.[158] However, the defendant then changed his mind and contended that he would not accept any refinancing loan, which led him to threaten the exercise of the 'foreclose' option on the NFT.[159] The claimant thereafter found the NFT for sale on Opensea, and sought a proprietary injunction against the defendant to prohibit the defendant from dealing in any way with the NFT.[160]

The nub of the case was therefore the question of whether NFTs could give rise to proprietary rights. There, the court noted that NFTs, when distilled to 'the base technology, are not just mere information, but rather, data encoded in a certain manner and securely stored on the blockchain ledger', and to characterise NFTs as mere information 'would ignore the unique relationship between the encoded data and the blockchain system which enables the transfer of this encoded data from one user to another in a secure, and verifiable fashion'.[161] The real objection to treating information as property, according to the court, depended on the 'function it is used for rather than the plain fact it is information'.[162] For NFTs, the information concerned was 'a string of computer code that does not provide any knowledge to those who have read it', which instead 'provides instructions to the computer under a system whereby the "owner" of the NFT has exclusive control over its transfer from his wallet to any other wallet'.[163] The court also observed that there had been growing support for 'deploying property concepts to protect digital assets' and cited various cases such as *Money-4 Ltd* for this proposition.[164]

The court also canvassed *AA v Persons Unknown*.[165] Whilst the English and New Zealand courts have accepted the *Ainsworth* criteria as the *prima facie* test for whether something may constitute property and approved of the UKJT statement to the effect that Fry LJ did not limit the scope of what could constitute property in *Colonial Bank v Whinney*, the court in *Janesh* instead cited Low's commentary in

[156] ibid [16].
[157] ibid [17].
[158] ibid [19].
[159] ibid [19]–[21].
[160] ibid [25].
[161] ibid [58].
[162] ibid.
[163] ibid.
[164] ibid [59].
[165] ibid [60]–[62].

criticism of the UKJT statement.[166] Indeed, the UKJT statement, heavily relied on by Bryan J in *AA* for the proposition that crypto assets could still constitute property, was viewed by Low as containing a lacuna.[167] This was because of the following portion of the statement:

> Thus, to the extent that the House of Lords [in *Colonial Bank*] agreed with Fry LJ on the classification issue, that seems to have been on the basis that the class of things in action could be extended to all intangible property (i.e. it was a residual class of all things not in possession) rather than on the basis that the class of intangible things property should be restricted to rights that could be claimed or enforced by action.[168]

Low viewed the preceding portion of the UKJT statement as an 'oxymoron', as the only way the statement could have 'ma[de] any sense [was] by disassociating the category of things in action in its first half from the narrow view of the enforceability of rights in the sense of Hohfeldian claim rights in its second'.[169] The Court in *Janesh* succinctly summarises Low's dissent in the following form. According to the UKJT, the House of Lords in *Colonial Bank* agreed with Fry LJ on the classification issue, seemingly on the view that the class of chose in action could be extended to all intangible property ('View A'), and not the view that the class of intangible property should be restricted to rights that could be claimed or enforced by action ('View B').[170] This was, however, paradoxical. This is because if a 'chose in action' (as expressed in View A) was referred to in the traditional sense (that is, rights or claims enforceable by action), this would render View A the equivalent of View B.[171] This, however, was not what was expressed in the UKJT's statement (as is reproduced above). Indeed, the use of the word 'rather' in the statement seems to suggest that View A and View B stand for contrasting positions. In this regard, according to Low, the only way out of this rabbit hole was to read the 'chose of action' referred to in View A as not referring to rights that could be claimed or enforced by action. The natural inference of this play on logic was to render View A as going beyond 'mere rights enforceable by action'.[172] According to the court in *Janesh*, such a position might accord with the historical roots of property law in the round.

The reason for this conclusion is that in the past, the term 'chose in action' initially encompassed all rights which were enforceable by action, which included,

[166] *Janesh* (n 75) [63]; Kelvin FK Low, 'Bitcoins as Property: Welcome Clarity?' (2020) 136 Law Quarterly Review 345, 347–348.
[167] Low (n 166) 347–348.
[168] UK Jurisdiction Taskforce (n 105) para 76.
[169] Low (n 166) 348.
[170] *Janesh* (n 75) [65].
[171] *Janesh* (n 75) [66].
[172] ibid.

amongst other things, rights to a debt, or action on a contract.[173] However, choses in action were later extended to cover 'documents such as bonds, which evidenced or proved the existence of such rights of action'.[174] Subsequently, the ambit of what a chose in action constituted expanded, and the term consequently included instruments such as bills of lading, and even policies of insurance. In accepting policies of insurance—which in substance, were documents to title of what 'was essentially an incorporeal right to property'—as falling within this category, the stage was then set for the expansion of 'choses in action' to include other things which were 'even more obviously property of an incorporeal type'.[175] This included things such as patents and copyright.[176] In this regard, adopting an expansive interpretation of View A would likely accord with the historical roots of property law, and might therefore be preferable.

Notwithstanding this, the court recognised that the meaning of terms such as 'choses in action' or 'intangible property'—as is commonly used in judicial discourse—might not be entirely clear cut. Adopting Low's perspective (in expanding View A) would lead us to the position that both terms are co-extensive, which would open the gateway for the application of Fry LJ's '*tertium quid*' in answering the question of whether crypto assets were indeed property under the common law. The court noted that there was support for such a view. Indeed, insofar as the objection exists that crypto assets cannot constitute property because they are neither tangibles nor choses in action, there is authority in *Ruscoe* to suggest that this objection is but a red herring.[177] The court in *Janesh* observed that the most that this objection could reach was to say that cryptocurrencies would have to be instead classified as choses in action.[178] Moreover, it would be paradoxical for the law to acknowledges a simple debt as qualifying for proprietary status, but to deny the same status to crypto assets, when the latter has more proprietary features than the former.[179] Hence, there remains an overhanging question amongst courts as to the exact status of crypto assets, alongside the trend the law should develop towards.

The court in *Janesh* further acknowledged that the ongoing uncertainty surrounding this debate may have played a role in the widespread use of the *Ainsworth* criteria in determining whether crypto assets should be considered property. In most cases involving litigation on such assets, such as *Ruscoe*, there has remained a trend of counsel omitting to push the point that the common law 'only recognised two classes of personal property', with the consequence that crypto assets did not fall into either class.[180] The lack of dissent by counsel against the doctrinal

[173] Michael Bridge, *Personal Property Law* (4th edn, Oxford University Press 2015) 229.
[174] WS Holdsworth, 'The History of the Treatment of "Choses" in Action by the Common Law' (1920) 33 Harvard Law Review 997.
[175] ibid 998.
[176] *Janesh* (n 75) [64].
[177] *Ruscoe* (n 124) [123].
[178] *Janesh* (n 75) [67].
[179] *Ruscoe* (n 124) [124].
[180] *Janesh* (n 75) [68].

foundations of property law as it relates to crypto assets has therefore led to courts adopting *Ainsworth* as a useful framework in judging this question during the interim.[181]

However, the *Ainsworth* test is not without its flaws, as acknowledged by the court in *Janesh*. Low has argued that the *Ainsworth* test 'mixes up the various meanings which common lawyers give to the word property', as what may be the subject matter of a trust, or the subject matter of a proprietary injunction, is much wider than what the *Ainsworth* test encompasses.[182] In other words, Low suggests that the *Ainsworth* test is overly restrictive and may not be the most appropriate criterion in determining what ought to be considered property under the law, and this view was tentatively acknowledged by the court in *Janesh*.[183] Despite these acknowledged flaws, the court, recognising the limitations of the present case and the tendency of common law courts to apply the *Ainsworth* test without question, then proceeded to use the *Ainsworth* test to determine whether NFTs constituted property, albeit with some hesitation owing to the possibility that a different conclusion could have been reached if more fuller submissions had been presented.[184]

On the first *Ainsworth* criterion, that is, definability, the court in *Janesh* held that this was easily fulfilled in the context of a NFT because 'metadata is central to an NFT, which distinguishes one NFT from another'.[185] On the second criterion, this being that the 'asset must have an owner being capable of being recognized by third parties', the court held that 'where NFTs are concerned, the presumptive owner would be whoever controls the wallet which is linked to the NFT', and thus excludability is achieved because one cannot deal with the NFT 'without the owner's private key'.[186] On the third criterion, that the 'right must be capable of assumption by third parties', the court held that the 'nature of the blockchain technology gives the owner the exclusive ability to transfer the NFT to another party, which underscores the "right" of the owner'; and that such NFTs are 'clearly the subject of active trading in the markets'.[187] On the final requirement, that there must be 'some degree of permanence or stability', the court considered that the 'NFT concerned has as much permanence and stability as money in bank accounts, which nowadays exist in the form of ledger entries and not cold hard cash'.[188]

Two lessons emerge in the wake of *Janesh*. First, it might be said that the utility of the *Ainsworth* test has finally been questioned by a common law court, and it will be interesting to see whether this test should continue holding water as the law develops. Second, it seems that NFTs are clearly able to satisfy the *Ainsworth*

[181] ibid.
[182] Low (n 166) 348–349
[183] *Janesh* (n 75) [69].
[184] ibid.
[185] ibid.
[186] ibid [70].
[187] ibid [71].
[188] ibid [72].

criteria, and this analysis will likely be relied upon by future courts when deciding the question of whether NFTs (and other digital assets) constitute property.

G. A PRELIMINARY CONCLUSION

The cases discussed above show that common law courts have also found themselves in a quandary. In the absence of express policy guidance, the *Ainsworth* test has been forcibly applied to address the domain of digital assets. Although helpful as a starting point, the *Ainsworth* criteria cease to hold weight when plunged into the murky depths of edge cases as might be common in digital assets. *Janesh*, however, suggests that there might indeed be a case for the development of another class of property beyond *Colonial Bank*'s antiquated bifurcation. Moreover, the Hohfeldian bundle of rights theory and its associated policy-based reasoning appears to be gaining ground within the juridical discourse, with courts recognising the potential implications that digital assets could have on wider society. In Section IV, this article will argue that the proper way forward is not the *Ainsworth* test, but through the proposal raised by the Law Commission, subject to certain tweaks.

IV. THE WAY FORWARD

As Low argues, the problem with the *Ainsworth* criteria is that the test 'mixes up the various meanings by which common lawyers use the term "property"'.[189] What qualifies as property in one context may not qualify as property in another. For example, Low helpfully illustrates that in *Ainsworth*, the criterion was used to deny proprietary status of an *in rem* right a 'deserted wife's right to absolve the bank of liability'.[190] In *B2C2*, the question at hand was whether 'cryptoassets were sufficiently property so as to be the subject matter of a trust'; but the use of property in this case was clearly different, because '*in personam* contractual rights may also be held on trust'.[191] Likewise, in *AA v Persons Unknown*, the proprietary injunctions were actually '*in personam* debt claims against a bank' at common law.[192]

The *Ainsworth* test is therefore rendered substantively hollow, and *Janesh* underscores the need to develop a new test for property that is suitable in the context of cryptocurrencies and NFTs. This section proceeds on this basis by examining the literature surrounding the present state of the law and suggests a way forward out of the current uncertainty.

Most academics have expressed a general intuition that a property rule should apply, even though the exact nature of crypto assets remains ambiguous. According to Fox, the subsisting property law framework can and should apply to

[189] Low (n 166) 349.
[190] ibid.
[191] ibid.
[192] Ibid.

crypto tokens, by segmenting crypto tokens as a special subset of intangible assets.[193] Even though a transaction might not be reversible on the blockchain system, traditional property principles should apply to allow tokens which were stolen or fraudulently transferred to be recovered, even if the blockchain system might not indicate whether said transaction is otherwise lawful.[194] Ng backs this up, arguing that in the context of theft, 'there is every reason to characterize the issue as proprietary'.[195] A property principle would most certainly serve as a steady hand to guide the law forward in this area, though one may question if *Ainsworth* should be the chosen one.

Notwithstanding, various scholars have expressed caution against recognizing crypto assets as property. Hewitt rightly points out that there remains the risk of 'blanket liens'— according to her, should banks gain an interest in all the proprietary rights held by a business, the moment Bitcoin is transferred to said business, the lien would then apply automatically, hence hindering liquidity.[196] But although this might create problems in relation to insolvency law, Sarra and Gullifer argue that Bitcoin should still be viewed as property notwithstanding the underlying difficulties, for the very reason that a crypto asset is an asset which has value.[197]

The fundamental question is therefore as follows: what exactly is property? Babie, Brown, Giancaspro, and Catterwell address this question accurately. Citing Ziff, they argue that the question might be answered in a bifurcated manner.[198] The first consists of an 'attributes approach', which mandates a court to locate an 'external indicator' that property does exist in the item in consideration by attempting to draw analogies between the novel case at hand and previously decided cases.[199] Such an approach, however, assumes that property as a static concept, an assumption which sits uneasily with technological developments. The better view is Ziff's 'functional approach'. According to Ziff, a judge must always remember that property is about a relationship consisting of the legal rights of use, excludability, and alienability.[200] The nub of the inquiry should instead focus on whether the relevant relationship exists at that point in time in respect of the thing or asset

[193] David Fox, 'Cryptocurrencies in the Common Law of Property' in David Fox and Sarah Green (eds), *Cryptocurrencies in Public and Private Law* (Oxford University Press 2019).
[194] ibid.
[195] Michael Ng, 'Choice of Law for Property Issues Regarding Bitcoin Under English Law', (2019) 15 Journal of Private international Law 315, 322.
[196] Evan Hewitt, 'Bringing Continuity to Cryptocurrency: Commercial Law as a Guide to the Asset Categorization of Bitcoin' (2016) 39 Seattle University Law Review 619, 629–630.
[197] Janis Sarra and Louise Gullifer, 'Crypto-claimants and Bitcoin Bankruptcy: Challenges for Recognition and Realization' (2019) 28 International Insolvency Review 242, 251.
[198] Paul Babie and others, 'Cryptocurrency and the Property Question' (*Faculty of Law Blogs: University of Oxford*, 12 May 2020) <https://blogs.law.ox.ac.uk/research-and-subject-groups/property-law/blog/2020/05/cryptocurrency-and-property-question> accessed 19 March 2023. See also Bruce Ziff, *Principles of Property Law* (7th edn, Carswell 2018).
[199] Babie and others (n 198).
[200] ibid.

in question. In the words of Ziff, this involves the consideration of how 'property, as a tool of social life, should be used'.[201]

This suggestion brings us full circle back to the Law Commission's proposal in Section II. There is a case for a third category of property, termed 'data objects' (as the Law Commission calls it). This third category of 'data objects', diverging from the Law Commission's proposal, should be defined by the following criteria: (a) it is composed of data represented in computer code; (b) it exists independently of persons and exists independently of the legal system; and (c) it is rivalrous. Such a definition incorporates digital assets into a proper Hohfeldian model, and correctly recognises the relationship that digital assets enjoy with society at large. There are various entry points for further research in this regard. Scholars should look towards how this third category might be expanded to accommodate more digital assets. This will allow other digital assets, such as digital files or domain names to be properly covered. Next, scholars might consider looking into how the tort of conversion might apply beyond that of crypto tokens, particularly given the fact that courts in other jurisdictions have held that digital assets can indeed be the subject of a conversion claim. Finally, scholars should investigate the doctrines of tracing and following. It is argued that the doctrine of following is the proper approach in this area, but the question remains as to whether, and if so how, the innocent purchaser defence might operate in this context.

V. Conclusion

In Section II, this article has evaluated the Law Commission's proposal for a third category of property. It has argued that the Law Commission's current proposal could be made more nuanced. It is inapplicable to digital assets other than cryptocurrency, legal uncertainty continues remains in relation to such assets, particularly in the context of cloud storage and other intangibles. Further, this article has argued that the Law Commission's criteria for 'data objects'—particularly its definition of data— and associated legal remedies could be further reworked. To build on the Law Commission's proposal, Section III of this article has explored how common law jurisdictions have treated digital assets in the context of the law of property, particularly through zooming into the policy set-up and the juridical technological discourse that has occurred to date. The conclusion, in Section IV, is that the Law Commission has presented a commendable proposal, though some tweaks are needed, particularly in the context of the test for 'data objects' and its associated remedies.

This paper proposes that the third category of property should be defined by the three-pronged test as set out earlier, though tweaked in terms of the type of data concerned (with the requirement being that it should be represented by computer code). Such a third category of property finds further support in cases across common law jurisdictions, with the recent case of *Janesh* calling into question

[201] ibid.

the future applicability of the *Ainsworth* test. Indeed, developing the law in this direction, and away from *Ainsworth*, would ground the future development of digital assets firmly within the Hohfeldian 'bundle of rights' theory, thereby reorienting the law of property in accord with policy and reality.

Cambridge Law Review (2023) Vol 8, Issue 1, 72–103

A Question Not for the Courts to Answer: A Thematic Analysis of the Recent Jurisprudence of the German Federal Constitutional Court and the Austrian Constitutional Court through the Lens of the Political Question Doctrine

CLARA VALERIA KAMMERINGER*

ABSTRACT

This article discusses how the German Federal Constitutional Court and the Austrian Constitutional Court have addressed political questions submitted for constitutional review in their jurisprudence from 2017 to 2021. 'Political questions' mainly concern the fields of foreign policy and security, the rules governing the democratic process, core political controversies, and possibly fundamental rights claims. Five themes are identified to discuss the Courts' practices: (a) discussions of legislative margins of appreciation; (b) references to external sources; (c) the offering of constitutionally conforming interpretations and guidelines; (d) the application of holistic policy considerations; and (e) discussions of the relationship between the Courts, and the legislature and the executive. The respective approaches are evaluated by reference to concerns relating to the separation of powers, checks and balances, fundamental rights, and judicial prudence. The evaluation of these practices yielded mixed results, with all practices having advantages and disadvantages. To improve the Courts' approaches, the paper outlines a political question doctrine similar to the one developed in US constitutional law. The political question doctrine renders certain political questions non-justiciable, based on their textual commitment to another branch of government or for prudential reasons. The doctrine proposed for Germany and Austria includes judicial restraint for issues being debated in parliament, and possibly the option for

* BA (Maastricht University). I would like to thank Dr Prashant Sabharwal for supervising this research, and offering valuable advice and encouraging words, as well as all who have supported and encouraged me throughout the process of writing this article. I would also like to thank the University College Maastricht and its academic community for allowing me to embark on this journey.

the Courts to initiate political discussions or popular consultations. However, if the Courts observe a risk of serious fundamental rights infringements, they should be able to issue a decision remedying the violation, despite the initial applicability of the doctrine.

Keywords: comparative constitutional law, political question doctrine, German Federal Constitutional Court, Austrian Constitutional Court, thematic case law analysis

I. Introduction

'The law encompasses any action... The fact that an issue is "strictly political" does not change the fact that such an issue is also "a legal issue".'
Aharon Barak[1]

'[Modern human rights law] *transforms controversial political issues into questions of law for the courts. In this way, it takes critical decision-making powers out of the political process.*'
Jonathan Sumption[2]

Constitutional courts are seen as protectors of human rights, guarantors of the rule of law, and arbiters of constitutional disputes. However, their decisions are not without controversy, and scholars debate their proper functions and powers. Constitutional courts are 'established, independent organ[s] of the state whose central purpose is to defend the normative superiority of the constitutional law within the juridical order'.[3] Constitutional review is one of the key means by which constitutional courts perform this task. It is defined as 'the power of judicial bodies to set aside ordinary legislative or administrative acts if judges conclude that they conflict with the constitution'.[4] In the centralised system of constitutional review, constitutional courts are the only courts that can exercise constitutional review.[5]

Constitutional courts were conceived by Hans Kelsen when framing the Constitution of the First Republic of Austria (1920–34).[6] After the Second World War, the German Federal Constitutional Court was established based on the Austrian model. After the collapse of the authoritarian regimes in Southern and Eastern Europe in the 1970s and 1990s, the majority of the newly founded nations adopted the Austro-German constitutionalist approach. The underlying assumption was that constitutional courts would enable a robust system of rights

[1] HCJ 910/86 *Major (Res) Yehuda Ressler v Minister of Defence* (Israeli High Court of Justice, 12 June 1988) [477].
[2] Jonathan Sumption, *Trials of the State: Law and the Decline of Politics* (Profile Books 2019) 60.
[3] Alec Stone Sweet, 'Constitutional Courts' in Michel Rosenfeld and András Sajó (eds), *The Oxford Handbook of Comparative Constitutional Law* (Oxford University Press 2012) 817.
[4] Georg Vanberg, *The Politics of Constitutional Review in Germany* (Cambridge University Press 2005) 1.
[5] Stone Sweet (n 3) 817–818.
[6] ibid 817–819.

protection, a prerequisite for democracies. However, it is important to recognise that constitutional courts fulfil a legislative function through the exercise of constitutional review, as they have the authority to strike down legislation.[7]

In recent decades, many scholars have observed a 'judicialisation of politics',[8] that is, a global trend towards shifting powers from representative institutions to the judiciary.[9] In his book *Governing with Judges*, Alec Stone Sweet argues that policymaking has been judicialised 'by an ever-expanding web of constitutional constraints' that allows '[c]onstitutional judges [to] routinely intervene in the legislative process'.[10] Some argue that the powers of constitutional courts should be limited by curtailing the justiciability of 'political questions'.[11] This is because there may not be an appropriate legal basis for answering these questions, the executive or legislative branches may have been explicitly tasked with answering them, or the court may suffer negative consequences, such as a decline in its legitimacy, as a result of providing an answer.[12] Therefore, US constitutional law has long developed a political question doctrine which renders certain political questions non-justiciable.[13] However, its precise form is not clear, and its application is inconsistent.[14]

This article aims to examine: (a) how the German Federal Constitutional Court and the Austrian Constitutional Court have addressed political questions submitted for constitutional review between 2017 and 2021; and (b) whether the approaches of the respective courts are desirable when considering aspects such as the separation of powers, checks and balances, fundamental rights, and concerns of judicial prudence. Section II discusses the political question doctrine and defines what constitutes a political question for the purposes of this article. Section III explains the methodology used to select and analyse the Courts' decisions. Section IV provides a brief overview of the history of the German Federal Constitutional Court and the Austrian Constitutional Court and traces the development of the jurisprudence of these courts in relation to political questions. Section V

[7] ibid 819.

[8] See for example Ran Hirschl, 'The Judicialisation of Politics' in Robert E Goodin (ed), *The Oxford Handbook of Political Science* (Oxford University Press 2011) 255; C Neal Tate and Torbjörn Vallinder, 'The Global Expansion of Judicial Power: The Judicialisation of Politics' in C Neal Tate and Torbjörn Vallinder (eds), *The Global Expansion of Judicial Power* (New York University Press 1995) 5; Alec Stone Sweet, *Governing with Judges: Constitutional Politics in Europe* (Oxford University Press 2000) 1.

[9] Ran Hirschl, 'The New Constitutionalism and the Judicialization of Pure Politics Worldwide' (2006) 75 Fordham Law Review 721.

[10] Stone Sweet (n 8) 1.

[11] See for example J Peter Mulhern, 'In Defense of the Political Question Doctrine' (1988) 137 University of Pennsylvania Law Review 97, 175–176; Rachel E Barkow, 'More Supreme than Court? The Fall of the Political Question Doctrine and the Rise of Judicial Supremacy' (2002) 102 Columbia Law Review 237, 330–331.

[12] Zachary Baron Shemtob, 'The Political Question Doctrines: Zivotofsky v. Clinton and Getting beyond the Textual-Prudential Paradigm' (2016) 104 Georgetown Law Journal 1001, 1013, 1017.

[13] Martin H Redish, 'Judicial Review and the "Political Question"' (1984–85) 79 Northwestern University Law Review 1031, 1031.

[14] Alexandra Mercescu and Sorina Doroga, 'Should the European Court of Justice Develop a Political Question Doctrine?' (2021) 7 International Comparative Jurisprudence 15, 21.

answers the first question by identifying five themes of practice that exemplify the approach taken by the Courts in addressing political questions in the past five years. Furthermore, it presents an analysis of the themes according to the evaluative criteria so to answer the second question. Finally, Section VI sets out a possible political question doctrine for Germany and Austria.

II. Theoretical Framework

A. The political question doctrine

(i) The Doctrine's Development in the US Supreme Court

The political question doctrine has primarily developed in the case law of the US Supreme Court. It provides that if a 'political question' is brought before a court, the court ought not to answer the question as it is a non-justiciable matter. It is argued that the doctrine was established in *Marbury v Madison*.[15] In *Marbury*, Chief Justice Marshall held that '[t]he province of the court [was solely] to decide on the rights of individuals, not to enquire [how the other branches] perform duties in which they have a discretion'.[16] Thus, '[q]uestions, in their nature political, or which are, by the constitution and laws, submitted to the executive, can never be answered in this court',[17] but 'the decision of the executive is conclusive'.[18] After one and a half centuries, the US Supreme Court in *Baker v Carr* formulated more coherent guidelines on when courts should invoke the doctrine.[19] Justice Brennan identified six factors that indicate a political question:

1. 'A textually demonstrable constitutional commitment of the issue to a coordinate political department';
2. 'A lack of judicially discoverable and manageable standards for resolving it';
3. 'The impossibility of deciding without an initial policy determination of a kind clearly for nonjudicial discretion';
4. 'The impossibility of a court's undertaking independent resolution without expressing lack of the respect due coordinate branches of government';
5. 'An unusual need for unquestioning adherence to a political decision already made'; and

[15] *Marbury v Madison* 5 US 137 (1803).
[16] ibid [170].
[17] ibid.
[18] ibid [166].
[19] *Baker v Carr* 369 US 186 (1962).

6. 'The potentiality of embarrassment from multifarious pronouncements by various departments on one question'.[20]

However, the subsequent case law of the US Supreme Court has not been consistent and has not, despite *Baker*, brought forward a clear definition of the doctrine. In 2012 and 2015, the Supreme Court created further uncertainty surrounding the political question doctrine through its decisions in *Zivotofsky v Clinton* (*Zivotofsky I*)[21] and *Zivotofsky v Kerry* (*Zivotofsky II*)[22]. In *Zivotofsky I*, the Supreme Court narrowed the political question doctrine, focusing only on the first two *Baker* criteria.[23] In *Zivotofsky II*, the Supreme Court expressed broad and conclusive recognition of exclusive executive power. Thus, even if the political question doctrine is avoided, it does not necessarily lead to greater scrutiny by the Supreme Court. The Supreme Court might still recognise broad executive or congressional power that does not merely remove their acts from judicial review, but also cuts off all scrutiny, thereby constitutionally validating such acts.

The lack of clarity is also reflected in academic debate. Scholarly opinions analysing the US Supreme Court's case law regarding the political question doctrine span from assertions that the doctrine is in decline or has ceased to exist,[24] to views that it is growing,[25] to assertions that, substantively, it has never existed.[26]

(ii) Different Formulations of the Doctrine

Scholarly disagreement extends not only to the current state of the doctrine, but also its formulation and implications. There are multiple formulations of the political question doctrine discussed by legal scholars. Most of them can be categorised as either a textual understanding of the doctrine or a prudential version. According to the textual (or classical) version of the doctrine, a court ought not to make decisions that are explicitly committed by the constitution to another branch of government.[27] The prudential version renders questions non-justiciable if they are not fit for principled judicial decision making because of a lack of institutional competence, or if their answering would lead to political backlash or loss of confidence into the court by the people.[28]

[20] ibid [217], [222].
[21] *Zivotofsky v Clinton* 566 US 189 (2012).
[22] *Zivotofsky v Kerry* 576 US 1 (2015).
[23] Harland Grant Cohen, 'A Politics-Reinforcing Political Question Doctrine' (2017) 49 Arizona State Law Journal 1, 22.
[24] Barkow (n 11) 263ff; Mark Tushnet, 'Law and Prudence in the Law of Justiciability: The Transformation and Disappearance of the Political Question Doctrine' (2002) 80 North Carolina Law Review 1203, 1208–1209.
[25] Margit Cohn, 'Form, Formula and Constitutional Ethos: The Political Question/Justiciability Doctrine in Three Common Law Systems' (2011) 59 The American Journal of Comparative Law 675, 687.
[26] Louis Henkin, 'Is There a "Political Question" Doctrine?' (1976) 85 Yale Law Journal 597, 600.
[27] ibid 1017; Redish (n 13) 1039–1040.
[28] Shemtob (n 12) 1013–1014; Redish (n 13) 1043–1044.

A further discussion concerns the effect of the doctrine. Most scholars argue that the political question doctrine precludes consideration of the merits of the relevant decision. Tara Leigh Grove argues that the meaning of the doctrine has significantly changed with the *Baker* decision.[29] Under the pre-*Baker* doctrine, which she calls the traditional doctrine, the court was required to treat as conclusive certain factual determinations made by the political branches when engaging in a consideration of the merits of the case.[30] According to her, a shift with *Baker* was brought about partially by the academic debate surrounding the doctrine, which construed it in a way that precludes justiciability.[31] However, this modern political question doctrine, as Grove construes it, is not a doctrine of judicial restraint; rather, it is one that follows the trend of judicial supremacy: the court has increased its powers as it now decides legal and factual issues itself, and determines whether, and to what extent, any other branch may be involved in constitutional decision making, often concluding that it is to be the single arbitrator. [32]

B. Political Questions

The definitions of what constitutes a 'political question' offered in the formulations of the doctrine provide some guidance but are not conclusive. The scholarship generally agrees that it is not possible to clearly distinguish between the legal and the political sphere.[33] Thus, it is hardly possible to define political questions. Discussions of the political question doctrine mostly revolve around certain policy areas. Typically, these are questions relating to matters of foreign policy and security,[34] or the rules governing the democratic process, such as decisions on the outcome of elections or election processes, access to public office, the (re-)distribution of political power, or the legality of political parties.[35] Another field is what Hirschl calls the realm of pure or 'mega' politics.[36] He understands this as 'core political controversies that define (and often divide) whole polities'.[37] The answer to these questions would 'define a polity's very *raison d'être*'.[38] Lastly, it has been argued that certain right-based claims might be considered as political

[29] Tara Leigh Grove, 'The Lost History of the Political Question Doctrine' (2015) 90 New York University Law Review 1908, 1913.
[30] ibid 1918, 1922–1924.
[31] ibid 1947, 1954–1956.
[32] ibid 1960, 1964, 1967.
[33] Hirschl (n 8) 257–258.
[34] John Harrison, 'The Political Question Doctrines' (2017) 67 American University Law Review 457, 460.
[35] Jesse H Choper, 'The Political Question Doctrine: Suggested Criteria' (2005) 54 Duke Law Journal 1457, 1479.
[36] Hirschl (n 8) 253.
[37] Hirschl (n 9) 727.
[38] ibid 728.

questions, given that they can be ethical questions for which legal texts provide no clear answer.[39]

III. Methodology

A. Comparative constitutional law

The academic field of comparative constitutional law only emerged after the Second World War,[40] and saw a significant growth at the end of the 20th century. One of the aims of the significantly older[41] field of the comparative study of statecraft has always been to design optimal institutions by comparing and understanding the advantages and disadvantages of existing models.[42] In past decades, comparative constitutional law has increasingly become an interdisciplinary field as constitutional questions often intertwine with other related fields including politics, sociology, and economics.[43]

This article analyses how Germany and Austria address political questions, considers the differences and similarities in their approaches, and views their practice through the lens of a doctrine developed in the United States. Simultaneously, the article aims to establish a better understanding of the practice of the German and the Austrian Constitutional Courts, assess it normatively, and provide recommendations for an improved approach.

B. Selection of court documents

(i) Jurisdictions

The German and the Austrian Constitutional Courts are quite similar when viewed from a structural or procedural perspective as well as in their constitutional review powers. However, their approaches to political questions are interesting to compare because of the Courts' significantly different historical underpinnings, their assigned roles within the state, and their different self-conceptions. The Austrian Federal Constitutional Law, enacted in 1919/20, is characterised by compromise and value neutrality.[44] The German Basic Law of 1948 was written to secure democracy as a form of governance and as a value system, and to protect itself

[39] Renata Uitz, 'Constitutional Courts in Central and Eastern Europe' (2007) 13 Judicial International 47, 58.

[40] Michel Rosenfeld and András Sajó, 'Introduction' in Michel Rosenfeld and András Sajó (eds), *The Oxford Handbook of Comparative Constitutional Law* (Oxford University Press 2012) 6.

[41] Rosenfeld and Sajó (n 40) 4–5; Rosalind Dixon and Tom Ginsburg, 'Introduction' in Tom Ginsburg and Rosalind Dixon (eds), *Comparative Constitutional Law* (Edward Elgar Publishing 2011) 1–2.

[42] Dixon and Ginsburg (n 41) 2.

[43] ibid 1.

[44] Tamara Ehs, 'Der VfGH als politischer Akteur. Konsequenzen eines Judikaturwandels?' (2015) 44(2) Austrian Journal of Political Science 15, 20.

from self-destruction.[45] According to scholars, the German Federal Constitutional Court tends to make general constitutional deductions and fundamental statements with a certain pathos, whereas the Austrian Constitutional Court is more focused on individual case decisions and is more formal.[46] As Michael Holoubek put it: the German Court is more of a *constitutional* court, the Austrian Court rather a constitutional *court*.[47]

(ii) Proceedings and Court Formation

Germany and Austria both follow the Kelsenian model of centralised constitutional adjudication,[48] where both Courts can review the constitutionality of laws.[49] Although the Austrian Constitutional Court can also review the lawfulness of ordinances,[50] these decisions are left out deliberately for matters of comparability. Thus, proceedings before the German Federal Constitutional Court initiated under articles 93(1)(2),[51] 93(1)(4a),[52] or 100(1)[53] of the German Basic Law and proceedings before the Austrian Constitutional Court initiated under article 140 of the Austrian Federal Constitutional Law[54] are considered in this article.

In terms of court formation, this article only considers decisions made by the two Senates of the German Federal Constitutional Court.[55] In Austria, it is the norm that the entire Court decides a case; only if the answer is sufficiently clear will a smaller panel respond.[56] Thus, only decisions by the full Court will be taken into consideration.

[45] Jürgen Jekewitz, 'Bundesverfassungsgericht und Gesetzgeber. Zu den Vorwirkungen von Existenz und Rechtsprechung des Bundesverfassungsgerichts in den Bereich der Gesetzgebung' (1980) 19 Der Staat 535, 538.

[46] Ehs (n 44) 21–22.

[47] Michael Holoubek, 'Wechselwirkungen zwischen österreichischer und deutscher Verfassungsrechtsprechung' in Detlef Merten (ed), *Verfassungsgerichtsbarkeit in Österreich und Deutschland* (Duncker & Humbolt 2008) 111.

[48] Christoph Hönnige, 'Verfassungsgerichte in den EU-Staaten: Wahlverfahren, Kompetenzen und Organisationsprinzipien' (2008) 6 Journal for Comparative Government and European Policy 524, 525.

[49] Germany: Basic Law for the Federal Republic of Germany 1949, arts 93(1)(2), 93(1)(4a), 100(1); Austria: Federal Constitutional Law 1930, art 140.

[50] Austria: Federal Constitutional Law 1930, art 139.

[51] Abstract review initiated by the Federal Government, a Land government, or one-fourth of the Members of the Bundestag.

[52] Individual complaint.

[53] Concrete review initiated by an ordinary court.

[54] In particular, art 140(1)–(3).

[55] Only the Senates are authorised to invalidate laws (Act on the Federal Constitutional Court, art 93c(1)) and Chambers can only grant constitutional complaints that are manifestly well-founded, and the issue has already been decided by the Federal Constitutional Court (Act on the Federal Constitutional Court, art 93c(1)). See also Bundesverfassungsgericht, 'Proceedings – Reaching a Decision' (*Bundesverfassungsgericht*) <www.bundesverfassungsgericht.de/EN/Verfahren/Der-Weg-zur-Entscheidung/der-weg-zur-entscheidung_node.html> accessed 9 April 2022.

[56] Constitutional Court Act 1953, art 7(1) holds that a quorum is given when the chairman and at least eight out of the twelve voting members are present. Article 7(2)(1) holds that the chairman and four voting members constitute a quorum if the legal matter is already sufficiently clarified by a prior decision of the

(iii) Timeframe

The timeframe of publication of the judgements analysed is 1 January 2017 through 31 December 2021. One aim of this article is to consider whether the approaches the Courts take are desirable. To make a meaningful assessment and to give recommendations for possible changes, it is most sensible to consider the Courts' recent decisions.

(iv) Judgement Selection

The advanced search functions of both Court databases were used.[57] In both Courts, decisions will be considered in German, because not all decisions are translated into English. In the database of the German Federal Constitutional Court, the three relevant types of proceedings—Abstract Judicial Review (BvF),[58] Specific Judicial Review of Statutes (BvL),[59] and Constitutional Complaint (BvR)[60]—were selected. In the Austrian Legal Information System under 'Verfassungsgerichtshof', the document type 'decision texts' (*Entscheidungstexte (TE)*) was selected and 'B-VG Art140' was entered as a search term.[61] A first sample was created by filtering for decisions concerning matters of foreign policy and security, rules governing the democratic process, and decisions involving core political controversies and controversial ethical questions often related to fundamental right claims.[62] The final sample was selected by conducting a preliminary analysis of these documents based on the definition of political questions and the evaluative criteria to be applied.[63]

C. Coding and Thematic Analysis

To establish an account of the practices employed by the Courts when approaching political questions the methods of coding and thematic analysis were used. Thematic analysis is a method for identifying and analysing patterns

Constitutional Court. See also Verfassungsgerichtshof Österreich, 'The Court's Bench and its Judicial Activity – Organisation of the Judicial Activity of the Court' (*VFGH: Verfassungsgerichtshof Österreich*) <www.vfgh.gv.at/verfassungsgerichtshof/organisation/the_courts_bench.en.html> accessed 9 April 2022.

[57] Decisions are published on the website of the Federal Constitutional Court <https://www.bundesverfassungsgericht.de/DE/Entscheidungen/Suche/suche_node.html> and in the Legal Information System of the Republic of Austria <https://www.ris.bka.gv.at/Vfgh/>.

[58] German Basic Law, art 93(1)(2); 13 results.

[59] Germany: Basic Law for the Federal Republic of Germany 1949, art 100(1). 55 results.

[60] Germany: Basic Law for the Federal Republic of Germany 1949, art 93(1)(4a). 1445 results.

[61] 638 results, the majority of which bore the relevant classificatory sign G of proceedings initiated under article 140 of the Federal Constitutional Law. Because it is not possible to search directly for the classificatory sign G the search was conducted for the relevant article in the Constitution. Decisions bearing another classificatory sign were filtered out by the researcher herself.

[62] Sample of 99 decisions.

[63] Final sample of 54 decisions.

(themes) within qualitative data.[64] A theme captures something important about the data in relation to the research question and represents a level of patterned response or meaning within the data. The research mostly followed an inductive approach, meaning that the themes emerged from the data themselves. To establish the themes, the documents were coded—that is, analysed—according to relevant features reappearing in the decisions. The themes of practices established based on the codes were subsequently evaluated according to the criteria set out below.

D. Evaluative criteria

The evaluative criteria have been chosen because they refer to classical critiques of constitutional review and constitutional courts' essential tasks. Prudential concerns has been chosen as an evaluative criterion because they relate to one of the formulations of the political questions doctrine and are interesting to consider in response to critiques of judicialisation.

(i) Separation of Powers

The principle of the separation of powers means that state power is divided between the three different branches of government.[65] The engagement of constitutional courts with political questions bears the risk of breaching the principle as they might answer questions designated for the other branches.[66] At the same time, judicial abstention draws the line between political and judicial power in favour of the political decision-maker.[67] This could also be contrary to the principle.[68]

(ii) Checks and Balances

The principle of checks and balances holds that the different branches of government are responsible for controlling each other and ensuring each branch does not overstep its mandate.[69] It has been argued that with the political question doctrine, courts leave potentially unconstitutional governmental or legislative

[64] David E Gray, *Doing Research in the Real World* (4th edn, Sage Publications 2018) 692–694.
[65] Merriam-Webster, 'Separation of Powers' (*Merriam-Webster*) <www.merriam-webster.com/legal/separation%20of%20powers> accessed 12 June 2022.
[66] *Baker* (n 19) [210], [217]; Ariel L Bendor, 'Are There Any Limits to Justiciability: The Jurisprudential and Constitutional Controversy in Light of the Israeli and American Experience' (1997) 7 Indiana International & Comparative Law Review 311, 328.
[67] Harrison (n 35) 506.
[68] Jared P Cole, 'The Political Question Doctrine: Justiciability and the Separation of Powers' (Congressional Research Service 2014) <https://sgp.fas.org/crs/misc/R43834.pdf> 25 accessed 12 June 2022.
[69] 'Checks and Balances' (*Merriam-Webster Online*) <www.merriam-webster.com/dictionary/checks%20and%20balances> accessed 12 June 2022.

action unchecked.[70] Thus, judicial abstention might violate the principle, because the courts do not carry out their assigned task.[71]

(iii) Fundamental Rights Protection

The development of fundamental rights and their increasing importance made it possible for constitutional courts to deal with politically sensitive questions.[72] It has been argued that certain fundamental right claims lead courts to engage in possibly inappropriate policymaking.[73] Nonetheless, the protection of fundamental rights is one of the key tasks of constitutional courts and must be preserved.

(iv) Concerns of judicial prudence

Some authors have grounded the political question doctrine on prudential concerns. They refer to instances where courts lack institutional competence because there is no clear legal basis, and where interference could cause backlash from the other branches and decrease citizens' trust in the court.[74] Arguably, courts should act prudently to remain legitimate.

E. LIMITATIONS

This research is limited by the ambiguity of the political question doctrine and the term 'political question'. This influences the selection of decisions considered in this research. Because objective criteria do not exist, which decisions qualify for consideration under the political question doctrine depends on a degree of subjective judgment by the author. Furthermore, because the selection was mostly conducted by the author herself, the risk of missing a case that would have met the selection criteria remains. Additionally, the cases considered represent instances where the Courts did issue a decision. The Courts may have rejected questions for being non-legal, but this is not represented in this research as rejections are not considered in the sample. Lastly, the inductive approach chosen for the thematic analysis leaves open the possibility that a potentially relevant theme was missed or that aspects of decisions could be categorised differently.

[70] Redish (n 13) 1054.
[71] Martin H Redish, 'Abstention, Separation of Powers, and the Limits of the Judicial Function' (1984) 94 Yale Law Journal 71, 74.
[72] Uitz (n 40) 58.
[73] See for example Hirschl (n 9) 276; Uitz (n 40) 54.
[74] See for example Shemtob (n 12) 1013.

IV. Development of Constitutional Review in Germany and Austria

Judicial decisions usually follow case precedents. Thus, before discussing the analysis of the case law from the past five years, a brief overview of the development of constitutional review at the two Courts and their jurisprudence concerning political questions is appropriate. The aim is to provide an understanding of how the two Courts have developed and acted, so to put their more recent decisions in context.

A. Developments in Germany

It has been argued that the German Federal Constitutional Court entered its most formative years immediately after it was set up in 1951,[75] when the Court laid the foundations of its fundamental rights jurisprudence.[76] When Willy Brandt became Chancellor in 1969, the Court suddenly found itself in the conservative role of the guardian of constitutional 'values' against the politics of 'progress'.[77] During the 1970s and 1980s, the Court arguably withdrew from the high political stage by assuming a more doctrinal rhetoric.[78] However, on multiple occasions, the Court has been criticised for arbitrary decisions,[79] taking political sides, or abandoning a basic political consensus.[80] Scholars now see increasing indications that the Court has passed its zenith.[81] This is explained by the stability and prosperity of the Federal Republic over the past decades, and the Court's loss of charisma through routinisation.

B. Developments in Austria

The development of constitutional adjudication in Austria can be described in different phases.[82] In the First Republic, the Constitutional Court was hesitant

[75] Florian Meinel, 'The Constitutional Miracle on the Rhine: Towards a History of West German Constitutionalism and the Federal Constitutional Court' (2016) 14 International Journal of Constitutional Law 277, 283.
[76] ibid 284.
[77] ibid 286.
[78] ibid 288.
[79] Dietmar Willoweit, 'Rechtsprechung und Staatsverfassung: Zur Geschichte und Gegenwart einer ambivalenten Beziehung' (2016) 71 JuristenZeitung 429, 432.
[80] Christoph Schönberger, 'Karlsruhe' in Matthias Jestaedt and others (eds), *The German Federal Constitutional Court: The Court Without Limits* (Oxford University Press 2020) 23.
[81] ibid 27.
[82] Heinz Schäffer, 'Verfassungsgericht und Gesetzgebung' in W Berka, H Schäffer, and J Werndl (eds), *Staat – Verfassung – Verwaltung* (Springer 1998) as cited in Ehs (n 44) 19; see also Theo Öhlinger, 'Die Verfassungsgerichtsbarkeit in Österreich – Der Wandel von Funktion und Methode in einer neunzigjährigen Geschichte' in Christian Boulanger and Michael Wrase (eds), *Die Politik des Verfassungsrechts* (Nomos 2013).

towards scrutinising legislation based on fundamental rights. After World War II, the Court continued to practice judicial self-restraint, although it tentatively began to show the legislature its limits, though the legislator's realm of policymaking remained untouched. Only since the 1980s did the Court start to interpret fundamental rights more substantively, when it became in part judicially activist and started determining the limits of legal intervention by applying the principle of proportionality. However, in the past ten years, the Court seems to have become less activist again.[83] While maintaining its role as the guardian of the constitution and fundamental rights, the Court has ceased to further develop its principles.

V. Thematic Analysis and Evaluation

The literature suggests that both Courts answer at least some political questions.[84] The research conducted confirmed this. Although not all of the decisions now discussed as potential cases for the doctrine might be considered as such by a court, they concern the fields defined pertaining to political questions as set out in the Theoretical Framework above,[85] and are thus deemed relevant to the present analysis. The analysis and evaluation that follow discuss how the Courts approach these questions on an exemplary basis.

A. Legislative margin of appreciation

(i) Theme: Legislative Margin of Appreciation

Both Courts refer to the concept of a legislative margin of appreciation and use it quite similarly, in that the legislator has some freedom in making policy decisions. If contested legislation falls within the legislator's margin of appreciation, it will not be found unconstitutional. This concept mirrors a formulation of the political question doctrine whereby the court accepts certain decisions of the legislative branch and treats them as facts.[86]

Both Courts recognise that the legislator has a margin of appreciation when having to make complex decisions,[87] these being decisions involving the balancing

[83] Konrad Lachmayer, 'Formalism and Judicial Self-Restraint as Tools Against Populism? Considerations to Recent Developments of the Austria Constitutional Court' in Fruzsina Gárdos-Orosz and Zoltán Szente (eds), *Populist Challenges to Constitutional Interpretation in Europe and Beyond* (Pre-print version, Routledge 2021) 15.

[84] See for example Thomas M Franck, *Political Questions/Judicial Answers* (Princeton University Press 1992) 108; Rudolf Streinz, 'The Role of the German Federal Constitutional Court: Law and Politics' (2014) 31 Ritsumeikan Law Review 95, 101; Lachmayer (n 83) 17; Öhlinger (n 82) 251.

[85] See Section II.B.

[86] Grove (n 29) 1924.

[87] BVerfG, Beschluss des Ersten Senats vom 19 November 2021 - 1 BvR 781/21, 1 BvR 889/21, 1 BvR 860/21, 1 BvR 854/21, 1 BvR 820/ 21, 1 BvR 805/21, 1 BvR 798/21 - Rn (1 - 306) [171].

of different rights,[88] multiple possible options to choose from,[89] or different concerns to be taken into consideration.[90] The German Federal Constitutional Court has further held that the legislator is allowed to make generalisations for reasons of predictability and simplicity of the law.[91] For example, when discussing the constitutionality of contact restrictions and curfews as a response to the COVID-19 pandemic, the German Federal Constitutional Court held that for measures concerning the pandemic, the legislator has a wide margin of appreciation because of the complexity and unforeseeability of the issue.[92] Thus, the Court found the measures to be overall constitutional.[93] Similarly, the Austrian Constitutional Court left a wide margin of appreciation to the legislator when countering the economic impact of the pandemic,[94] holding that the measures taken were constitutional.[95]

Both Courts, furthermore, recognise that the extent of the legislator's margin of appreciation depends on the severity of the infringement of fundamental rights.[96] Hence, there is a wider margin when there is no or only limited infringement. When discussing the prohibition of assisted suicide, the Austrian Constitutional Court held that there is no wide legislative margin of appreciation in this respect.[97] This is because the matter concerns an existential decision on the shaping of one's life and death and, thus, quite essentially the individual's right to self-determination. Similarly, the German Federal Constitutional Court held that the penalisation of commercialised assisted suicide, whilst following a legitimate aim and being as such a suitable instrument,[98] is not proportionate.[99] The regulation as discussed was found to completely empty the right to suicide as a manifestation of the right to a self-determined death in certain situations.

The German Federal Constitutional Court further grants the legislator a wide margin of appreciation when fulfilling its duty to protect.[100] Hence, the Court

[88] BVerfG, Urteil des Zweiten Senats vom 26 Februar 2020 - 2 BvR 2347/ 15, 2 BvR 2527/16, 2 BvR 2354/16, 2 BvR 1593/16, 2 BvR 1261/16, 2 BvR 651/16 - Rn (1 - 343) [224].
[89] BVerfG, Urteil des Zweiten Senats vom 19 September 2018 - 2 BvF 1/ 15, 2 BvF 2/15 - Rn (1 - 357) [171], [173], [284].
[90] VfGh 18.06.2019, G 150-151/2018-34, G 155/2018-32 [63], [64].
[91] BVerfG, Urteil des Ersten Senats vom 19 Dezember 2017 - 1 BvL 3/14, 1 BvL 4/14 - Rn (1 - 253) [187].
[92] BVerfG, Beschluss des Ersten Senats vom 19 November 2021 - 1 BvR 781/21, 1 BvR 889/21, 1 BvR 860/21, 1 BvR 854/21, 1 BvR 820/ 21, 1 BvR 805/21, 1 BvR 798/21 - Rn (1 - 306) [205].
[93] ibid [305].
[94] VfGH 14.07.2020, G 202/2020-20, V 408/2020-20* [118].
[95] ibid [135].
[96] VfGH 14.03.2017, G 164/2016-12 [108]; BVerfG, Beschluss des Zweiten Senats vom 23 Mai 2017 - 2 BvL 10/ 11, 2 BvL 28/14 - Rn (1 - 104) [97]; BVerfG, Beschluss des Ersten Senats vom 19 November 2021 - 1 BvR 781/21, 1 BvR 889/21, 1 BvR 860/21, 1 BvR 854/21, 1 BvR 820/ 21, 1 BvR 805/21, 1 BvR 798/21 - Rn (1 - 306) [204].
[97] VfGH 11.12.2020, G 139/2019-71 [83].
[98] ibid [227], [260].
[99] ibid [267].
[100] BVerfG, Beschluss des Ersten Senats vom 16. Dezember 2021 - 1 BvR 1541/20 - Rn (1 - 131) [99]; BVerfG, Beschluss des Ersten Senats vom 24. März 2021 - 1 BvR 2656/ 18, 1 BvR 96/20, 1 BvR 78/20, 1 BvR 288/20, 1 BvR 96/20, 1 BvR 78/20 - Rn. (1 - 270) [152].

will only find a violation if no measures were taken, or if the measures taken are obviously unsuitable, completely inadequate, or fall considerably short.[101] Both Courts allow for a margin of appreciation when assessing social needs,[102] or counteracting poverty, for example through minimum subsistence laws.[103] The German Federal Constitutional Court had to assess the constitutionality of the reduction of minimum subsistence payments in case of non-cooperation of the recipient in entering the labour market. In this context, it held that it is not the task of the Court to examine whether the legislator has chosen the most just, expedient, and reasonable solution to fulfil its tasks as this is for politics to try and achieve.[104] Rather, the Court must only ensure that the minimum subsistence level does not fall below a certain threshold and that the amount paid can be justified.[105] A similar sentiment was expressed by the Austrian Constitutional Court when discussing the merger of regional health insurance funds (*Gebietskrankenkassen*). In this context, the Court noted that it was not its appropriate role to review the political merit of the legislator's decisions.[106]

The Courts have referred to further policy areas where the legislator has a certain margin of appreciation, examples being when granting permanent residence and permissions to work for foreign nationals,[107] when creating taxes or charges and setting their rates,[108] or creating legislation for civil servants.[109] However, despite granting the legislator a margin of appreciation, the Courts have reviewed their actions according to standards of proportionality,[110] equality,[111] or other constitutional standards.[112]

[101] BVerfG, Beschluss des Ersten Senats vom 16. Dezember 2021 - 1 BvR 1541/20 - Rn (1 - 131) [98].
[102] VfGH 07.03.2018, G 136/2017-19 ua* [95].
[103] BVerfG, Urteil des Ersten Senats vom 5 November 2019 - 1 BvL 7/16 - Rn (1 - 225) [121].
[104] ibid [122].
[105] ibid [122].
[106] ibid [84].
[107] VfGH 04.10.2018, G 133/2018-12 [43]; VfGH 11.10.2017, G 56/2017-14, G 199/2017-8 [42].
[108] BVerfG, Beschluss des Zweiten Senats vom 13. April 2017 - 2 BvL 6/13 - Rn (1 - 45) [68]; BVerfG, Beschluss des Zweiten Senats vom 8 Dezember 2021 - 2 BvL 1/13 - Rn. (1 - 94), [55]; BVerfG, Urteil des Ersten Senats vom 18 Juli 2018 - 1 BvR 1675/16, 1 BvR 981/17, 1 BvR 836/17, 1 BvR 745/17 - Rn (1 - 157) [65], [71].
[109] BVerfG, Beschluss des Zweiten Senats vom 23 Mai 2017 - 2 BvL 10/ 11, 2 BvL 28/14 - Rn (1 - 104) [47].
[110] BVerfG, Urteil des Ersten Senats vom 18 Juli 2018 - 1 BvR 1675/16, 1 BvR 981/17, 1 BvR 836/17, 1 BvR 745/17 - Rn (1 - 157) [71], [106].
[111] VfGH 11.10.2017, G 56/2017-14, G 199/2017-8, [44], [47]; VfGH 07.03.2018, G 136/2017-19 ua* [104], [109]–[112], [114]; BVerfG, Beschluss des Zweiten Senats vom 8. Dezember 2021 - 2 BvL 1/13 - Rn (1 - 94) [53].
[112] BVerfG, Beschluss des Zweiten Senats vom 13 April 2017 - 2 BvL 6/13 - Rn (1 - 45) [128] (creation of taxes limited by types of taxes provided for in the constitution); BVerfG, Beschluss des Zweiten Senats vom 23 Mai 2017 - 2 BvL 10/ 11, 2 BvL 28/14 - Rn (1 - 104) [52] (take into consideration actual necessities and the development of financial and economic circumstances when setting deciding on the structure and the amount of remuneration for civil servants); BVerfG, Beschluss des Ersten Senats vom 16 Dezember 2021 - 1 BvR 1541/20 - Rn (1 - 131) [122], [130] (legislator did not take sufficient measures to fulfil their duty to protect).

(ii) Evaluation: Legislative Margin of Appreciation

In granting the legislator a margin of appreciation, the Courts adhere to the principle of the separation of powers to a large extent. They leave complex decisions to the legislative branch because of the legislator's democratically legitimated responsibility to decide conflicts between important interests, as noted by the German Federal Constitutional Court.[113] Furthermore, the Courts leave freedom to the legislator to undertake political reforms.

Nonetheless, the Courts also adhere to the principle of checks and balances, reviewing the merits of decisions even when the legislator has a margin of appreciation. By applying constitutional principles to these decisions, the Courts make it clear that it is still possible to overstep the margin they grant. Adherence to the principle of checks and balances is, however, slightly limited given that the Courts in some cases refrain from review or apply very low standards.

This links to the third criterion, the protection of fundamental rights. Despite the Courts' insistence on the protection of a right's core in any case, they do allow for some infringements by granting a wider margin for less severe infringements. Particularly, in granting a wide margin of appreciation in relation to the duty to protect and only mandating minimum standards for social needs, the German Federal Constitutional Court does not offer full protection of fundamental rights.

However, these and other concessions are most likely in line with prudential concerns. By refraining from reviewing the merits of political decisions and whether they are the best possible solutions, or only applying a limited review to some, the Courts limit themselves and avoid potential backlash.

B. Reference to external sources

(i) Theme: Reference to External Sources

Both Courts regularly refer to external sources, which include decisions from other jurisdictions, especially international courts, as well as reports and academic literature. It is not uncommon for courts to do so. Smyth, in a study of secondary source citation by the Australian High Court, mentions multiple reasons why courts refer to non-binding sources.[114] He notes the wish of a court to provide further justification for an interpretation or decision, and to refer to social sciences and other non-legal authorities to 'examine the "legislative fact" that underpins legal rules'.[115]

[113] BVerfG, Beschluss des Ersten Senats vom 19 November 2021 - 1 BvR 781/21, 1 BvR 889/21, 1 BvR 860/21, 1 BvR 854/21, 1 BvR 820/ 21, 1 BvR 805/21, 1 BvR 798/21 - Rn (1 - 306) [171].
[114] Russell Smyth, 'Other Than "Accepted Sources of Law"? A Quantitative Study of Secondary Source Citations in the High Court' (1999) 22 University of New South Wales Law Journal 19, 22–24.
[115] ibid 24.

Both Courts refer to case law from international courts, particularly the European Court of Human Rights (ECtHR) and the Court of Justice of the European Union (CJEU). Most references are made to the case law of the ECtHR, especially by the Austrian Constitutional Court. In multiple cases concerning security measures, the Austrian Constitutional Court has referred to the case law of the ECtHR, especially cases concerning article 8 (the right to respect for private and family life) of the European Convention on Human Rights (ECHR). Furthermore, both Courts refer to case law from the ECtHR on articles 2 (the right to life) and 8 ECHR when discussing the prohibition of assisted suicide.[116] In the cases concerning the official registration of non-binary genders, the German Federal Constitutional Court refers to case law from the CJEU,[117] while the Austrian Constitutional Court refers to the case law of the ECtHR.[118] In a case concerning the constitutionality of the European Public Sector Asset Purchasing Programme, the German Court engaged with the CJEU's case law concerning the matter.[119] However, the Court found that the CJEU did not deal with the case adequately,[120] and found the programme to be unconstitutional.[121]

Both Courts refer to international conventions, and may hold that certain measures are unconstitutional and also violate international laws.[122] On a number of occasions, the Austrian Constitutional Court referred to the International Convention on the Elimination of all Forms of Racial Discrimination.[123] At times, it also referred to the Geneva Refugee Convention[124] and the Convention on the Rights of the Child.[125] The German Federal Constitutional Court referred to violations of the Convention on the Rights of Persons with Disabilities once in the sample, as well as the International Covenant on Civil and Political Rights.[126] The court also made reference to General Comments by the Committee on Economic, Social, and Cultural Rights when discussing the prohibition and restriction of face-to-face teaching at general education schools to protect against infection during the COVID-19 pandemic.[127] Furthermore, when discussing measures for the

[116] VfGH 11.12.2020, G 139/2019-71 [67]–[71]; BVerfG, Urteil des Zweiten Senats vom 26 Februar 2020 - 2 BvR 2347/ 15, 2 BvR 2527/16, 2 BvR 2354/16, 2 BvR 1593/16, 2 BvR 1261/16, 2 BvR 651/16 - Rn (1 - 343) [305].

[117] BVerfG, Beschluss des Ersten Senats vom 10 Oktober 2017 - 1 BvR 2019/16 - Rn (1 - 69) [63].

[118] VfGH 15.06.2018, G 77/2018-9 [17], [18], [23].

[119] BVerfG, Urteil des Zweiten Senats vom 5 Mai 2020 - 2 BvR 859/15, 2 BvR 980/16, 2 BvR 2006/15, 2 BvR 1651/15 - Rn (1 - 237) [181]–[183].

[120] ibid [184]–[191].

[121] ibid [234].

[122] It should be noted that the Conventions referred to here have constitutional standing in Austria.

[123] VfGH 11.10.2017, G 56/2017-14, G 199/2017-8 [47]; VfGH 07.03.2018, G 136/2017-19 ua*, [131]; VfGH 12.12.2019, G 164/2019-25, G 171/2019-24 [131]; VfGH 26.06.2020 G 298/2019-11, G 117-121/2020-5 [31].

[124] VfGH 07.03.2018, G 136/2017-19 ua* [112], [114].

[125] VfGH 12.12.2019, G 164/2019-25, G 171/2019-24 [131].

[126] BVerfG, Beschluss des Ersten Senats vom 16 Dezember 2021 - 1 BvR 1541/20 - Rn (1 - 131) [100]–[107].

[127] BVerfG, Beschluss des Ersten Senats vom 19 November 2021 - 1 BvR 971/21, 1 BvR 1069/21 - Rn (1 - 222) [69], [172].

collection of intelligence, the Court highlighted letters from the UN High Commissioner for Human Rights, who had already criticised measures authorising the Federal Intelligence Service to engage in foreign-to-foreign telecommunications reconnaissance, the transmission of the information obtained to domestic and foreign agencies, and cooperation with foreign intelligence services.[128]

Both Courts further referred to the findings of expert reports. For example, in a case on the wearing of ideologically or religiously influenced clothing at school, the Austrian Constitutional Court referred to a report by the European Commission against Racism and Intolerance.[129] The report stated that the prohibition of wearing ideologically or religiously influenced clothing at school was unable to reach the legislator's goal of social integration. In a case concerning the constitutionality of measures taken in response to climate change, the German Court extensively discussed the findings of the reports by the Intergovernmental Penal on Climate Change (IPCC).[130]

The German Federal Constitutional Court on multiple occasions referred to secondary literature. For example, the Court referred to experts in statistics when assessing the constitutionality of the new method for establishing the national census.[131] In a case concerning the prohibition of commercialised assisted suicide, the Court referred to statistics of assisted suicide in other countries.[132] The Austrian Constitutional Court also referred to the prevailing view in literature to support its opinion[133] in holding that the prohibition of adoptive parenthood for non-married couples was unconstitutional.[134]

(ii) Evaluation: Reference to External Sources

Although not acting as positive policymakers, both Courts use external sources quite actively to engage in negative policymaking by holding that laws are unconstitutional. This can be seen as limiting the separation of powers. What is arguably more critical is the use of non-judicial sources as the bases of the Courts' argumentation. Although certain sources, such as the IPCC report, might be accepted as authoritative, references made to them might still be seen as policymaking rather than judicial decision-making.

Conversely, by considering multiple sources, the Courts ensure thorough scrutiny of the legislator. The practice ensures that the legislator not only adheres

[128] BVerfG, Urteil des Ersten Senats vom 19 Mai 2020 - 1 BvR 2835/17 - Rn (1 - 332) [240], [305], [325].
[129] VfGH 11.12.2020, G 4/2020-27 [144].
[130] BVerfG, Beschluss des Ersten Senats vom 24 März 2021 - 1 BvR 2656/ 18, 1 BvR 96/20, 1 BvR 78/20, 1 BvR 288/20, 1 BvR 96/20, 1 BvR 78/20 - Rn (1 - 270) [159]–[62], [178], [211], [215]–[223].
[131] BVerfG, Urteil des Zweiten Senats vom 19 September 2018 - 2 BvF 1/ 15, 2 BvF 2/15 - Rn (1 - 357) [167].
[132] BVerfG, Urteil des Zweiten Senats vom 26 Februar 2020 - 2 BvR 2347/ 15, 2 BvR 2527/16, 2 BvR 2354/16, 2 BvR 1593/16, 2 BvR 1261/16, 2 BvR 651/16 - Rn (1 - 343) [253]–[254].
[133] VfGH 06.12.2021, G 247/2021-12 [37].
[134] ibid [45].

to immediate national legal standards but also follows international and scholarly trends and is thus in line with the principle of checks and balances.

The extensive reference to and discussion of the case law of the ECtHR demonstrate a strong commitment to fundamental rights protection by both Courts. Interestingly, the German Federal Constitutional Court seems to refer to the ECtHR less often than the Austrian Constitutional Court. This might be because of fundamental rights provided by the German Constitution, whereas Austria's development of human rights law has always been interlinked with the ECHR.[135]

As to concerns of prudence, it can be positively noted that the Courts provide backing for their argumentation from other courts as well as non-judicial sources. However, as stated before, the Courts are also cognisant of the risks of engaging with broader external sources insofar as doing so may lead to the perception that they are no longer seen as engaging in a legal discussion, but explicitly engaging in policymaking through engaging with various sources, which might go beyond what is considered a legal discussion of the question and could be perceived negatively.

C. Constitutionally conforming interpretations and guidelines

(i) Theme: Constitutionally Conforming Interpretations and Guidelines

In response to non-conforming legislation, both Courts attempt to first provide a constitutionally compliant interpretation of the legislation, with the German Federal Constitutional Court doing so more often. Regardless, both Courts also provide recommendations or guidelines for the legislators on how to remedy the violation identified.

As to the provision of constitutionally compliant interpretations, one example is that the prohibition of certain associations in Germany was found to be constitutional.[136] In making such a finding, the German Federal Constitutional Court referred to the fact that even though a regulation lacked an explicit reservation of proportionality, the constitutional requirement of proportionality could be taken into account through interpretation.[137] The Austrian Constitutional Court, in a case concerning the possibility of officially registering one's gender as non-binary, likewise found it was possible to come to a constitutionally conforming interpretation of the current legislation.[138] This was because, in its view, the term 'gender' was broad enough to encompass non-binary genders, and thus the legislation

[135] See discussion on the matter in Section V.D.
[136] BVerfG, Beschluss des Ersten Senats vom 13 Juli 2018 - 1 BvR 1474/ 12, 1 BvR 57/14, 1 BvR 57/14, 1 BvR 670/13 - Rn (1 - 167) [96].
[137] ibid [118].
[138] VfGH 15.06.2018, G 77/2018-9 [45].

allowed people identifying as non-binary to be registered accordingly.[139] Furthermore, this broad interpretation recognises that someone's gender can be unidentified or changed.

The Courts sometimes also discuss why a constitutionally conforming interpretation is not possible. This is mostly because a constitutionally conforming interpretation would be contrary to the clear wording of the legislation[140] or the legislator's evident intent.[141] For example, the German Federal Constitutional Court found that that insofar as the gender registry referred clearly to 'male' and 'female',[142] it was unconstitutional.[143] Further, when discussing commercial assisted suicide, the German Federal Constitutional Court explicitly discussed the impossibility of a constitutionally conforming interpretation insofar as it would directly contradict the intent of the legislature on this area, and emphasised that any attempted constitutionally conforming interpretation would be tantamount to original judicial law-making and incompatible with the requirement of legal certainty.[144]

Where a provision has been declared constitutionally incompatible, both Courts sometimes prescribe requirements for the legislator to adhere to when creating new legislation on the matter discussed in a case. In holding that it has to be possible for someone to legally have access to assisted suicide, the Austrian Constitutional Court noted that the legislator had to consider that ways in which social and economic circumstances, and other circumstances outside of the person's control, can hamper a person's free self-determination.[145] Meanwhile, the German Federal Constitutional Court, in holding that the measures taken to combat climate change were insufficient, and therefore unconstitutional,[146] stated that one generation should not be allowed to consume large parts of the CO_2 budget under a comparatively mild reduction burden if this would leave a radical reduction burden to following generations and exposed their lives to severe losses of freedom.[147] Following this, the Court laid down certain requirements for the design of the CO_2 reduction scheme.[148]

[139] ibid [37], [38], [43].

[140] VfGH 08.03.2017, G 399/2016-8 [26]; BVerfG, Urteil des Ersten Senats vom 19 Dezember 2017 - 1 BvL 3/14, 1 BvL 4/14 - Rn (1 - 253) [151]; BVerfG, Beschluss des Zweiten Senats vom 8 Dezember 2021 - 2 BvL 1/13 - Rn (1 - 94) [90]; BVerfG, Beschluss des Ersten Senats vom 27 Mai 2020 - 1 BvR 1873/ 13, 1 BvR 2618/13 - Rn (1 - 275) [158].

[141] VfGH 11.12.2019, G 72-74/2019-48, G 181-182/2019-18 [224]; BVerfG, Beschluss des Zweiten Senats vom 8 Dezember 2021 - 2 BvL 1/13 - Rn (1 - 94) [90]; BVerfG, Beschluss des Ersten Senats vom 27 Mai 2020 - 1 BvR 1873/ 13, 1 BvR 2618/13 - Rn (1 - 275) [158].

[142] BVerfG, Beschluss des Ersten Senats vom 10 Oktober 2017 - 1 BvR 2019/16 - Rn (1 - 69) [42].

[143] ibid [65].

[144] BVerfG, Urteil des Zweiten Senats vom 26 Februar 2020 - 2 BvR 2347/ 15, 2 BvR 2527/16, 2 BvR 2354/16, 2 BvR 1593/16, 2 BvR 1261/16, 2 BvR 651/16 - Rn (1 - 343) [334].

[145] VfGH 11.12.2020, G 139/2019-71 [99]–[101].

[146] BVerfG, Beschluss des Ersten Senats vom 24. Mär̈z 2021 - 1 BvR 2656/ 18, 1 BvR 96/20, 1 BvR 78/20, 1 BvR 288/20, 1 BvR 96/20, 1 BvR 78/20 - Rn (1 - 270) [266].

[147] ibid [192].

[148] ibid [151]–[155].

(ii) Evaluation: Constitutionally Conforming Interpretations and Guidelines

Both Courts have stated that they will not interpret a provision in a way which is manifestly contrary to the legislator's intent as part of respecting the principle of the separation of powers. However, the constitutionally conforming interpretations provided by the Courts sometimes seem rather far-fetched, one example being when the Austrian Constitutional Court held that a same-sex parent adopting a child would step into the role of the parent of the other sex, thereby not replacing the biological parent.[149] Furthermore, they then provide an outright solution to cure the unconstitutional provision, implicitly circumscribing the legislator's scope to cure the defect themselves. Although constitutionally conforming interpretations do not inhibit the legislator from passing new legislation, issues with regards to judicial law-making and thus the separation of powers could arise. This is also the case when the Courts prescribe requirements as to how the legislator must remedy certain unconstitutionalities. Despite not legislating themselves, the Courts might still significantly limit the legislator's room for manoeuvre, especially as the requirements prescribed are often rather technical and might go beyond what is necessary for securing the legislation's constitutionality.[150]

On one hand, providing a constitutionally conforming interpretation is a sign of thorough scrutiny of the legislation and thus adherence to the principle of checks and balances. On the other hand, as has been seen, the Courts sometimes go rather far in their interpretation to avoid a finding of unconstitutionality and a repeal of the legislation. This could signal to legislators that the Courts, after a thorough review, will uphold the constitutionality of the legislation and thereby avoid damaging legislators' reputation. Thus, the legislator might perceive the Courts' acceptance of responsibility as a signal that they can give less consideration to the constitutionality and potential harm of legislation. Furthermore, a finding that a provision is constitutional may be misunderstood by the legislator, insofar as the court arrived at the conclusion via interpreting the provision in a way radically different from how the legislation is worded or previously understood.

Constitutionally conforming interpretations are mostly aimed at providing a solution that is more protective of fundamental rights. Similarly, the requirements for the legislator the Courts prescribe are meant to offer stronger protection of fundamental rights. However, they might not be the only or most desired rights-protecting solution, and can disincentivise the legislator from taking further action because it is seen to be unnecessary.

For prudential concerns, constitutionally conforming interpretations might make a court look modest because it does not strike down legislation as often.

[149] VfGH 03.102018, G 69/2018-9 [46]–[47].

[150] For example, the requirements for the design of the German CO_2 reduction scheme laid down by the German Federal Constitutional Court (BVerfG, Beschluss des Ersten Senats vom 24 Mäïrz 2021 - 1 BvR 2656/ 18, 1 BvR 96/20, 1 BvR 78/20, 1 BvR 288/20, 1 BvR 96/20, 1 BvR 78/20 - Rn (1 - 270) [192], [266].

However, upon closer consideration, the constitutionally conforming interpretation might amount to something close to policymaking, thus leading to potential political backlash.

D. Holistic policy considerations

(i) Theme: Holistic Policy Considerations

To discern the constitutionality of certain measures that could fall under the political question doctrine, both Courts resort to holistic policy considerations. This means that the Courts do not only consider the legislation in question but also its broader context. This can lead to both the declaration of constitutionality of a rather intrusive measure, or a finding of unconstitutionality in respect of measures which, viewed by themselves, seem constitutional. It appears that the Austrian Constitutional Court more often considers policies holistically, while the German Federal Constitutional Court tends to focus mainly on the legislation in question.

Both Courts refer to the 'fundamental characteristics' of the state or constitution. When discussing the expropriation of Hitler's birthplace, the Austrian Constitutional Court held that the uncompromising rejection of National Socialism was a fundamental characteristic of Austria.[151] Thus, it found the expropriation of the building to be constitutional for the purpose of eliminating the special symbolic power associated with the house through a profound architectural redesign.[152] When discussing the constitutionality of the establishment of an EU unified patent court, the German Federal Constitutional Court held that this would change the integration programme of the Treaty of Lisbon and create the possibility of a new type of unified jurisdiction in industrial property protection.[153] Further, it held that a transfer of jurisdictional tasks away from German courts would cause a change in the content of the Basic Law.[154] Hence, it held that the relevant law should not be passed.[155] It should have been treated as a constitutional amendment and would have needed a qualified majority in the Bundestag.[156]

As part of this holistic approach, the Austrian Constitutional Court also considers the constitutionality of the provision in question with reference to whether it is part of a broader policy package. When discussing the reduced time limit to appeal against return decisions, the Court found the measure to be unconstitutional because the public interest of clarifying the foreigner's status of residence as

[151] VfGH 30.06.2017, G 53/2017-23 [28].
[152] ibid [33].
[153] BVerfG, Beschluss des Zweiten Senats vom 1 Februar 2020 - 2 BvR 739/17 - Rn (1 - 168) [155].
[154] ibid [157].
[155] ibid [164].
[156] ibid [126].

soon as possible was not served by the measure taken.[157] This was the case because no other measures were taken that would accelerate the process in other stages of the proceedings, thereby rendering the measure incoherent.[158] Conversely, when discussing the absence of compensation for entry bans to businesses as a measure to curb COVID-19 infections, the Court did not find the legislation to be unconstitutional because, among other reasons, the measure had been taken as part of a comprehensive policy package with the overall aim of financial hardship of business owners.[159] Furthermore, the Austrian Constitutional Court considers how a matter is treated generally and refers to substantively similar situations. For example, when discussing same-sex marriage, the Court held that civil partnerships had been created for same-sex couples and that both marriage and civil partnerships signified an equal partnership and institutionalised a strong connection,[160] and that the two had been largely treated the same.[161] It therefore concluded that the unequal treatment of heterosexual and homosexual couples in marriage could no longer be upheld.[162]

Both Courts review measures in relation to their overarching goals. If this goal is not met by the measure, they find them to be unconstitutional even if the measure, by itself, would not be such. When discussing the constitutionality of reducing the minimum subsistence payments for non-cooperation, the German Federal Constitutional Court found that if the legislator pursued the legitimate goal of helping people avoid or overcome their own need for assistance, punitive measures to encourage such must be proportionate,[163] which was not the case.[164] The Austrian Constitutional Court, similarly, discussed a Viennese minimum income scheme which was held unconstitutional because it failed to achieve its actual purpose, namely the elimination of existing hardship.[165]

Lastly, as part of the Courts' endeavour to discern relevant policy considerations, they may have recourse to the history of certain laws when determining their meaning or proper interpretation. In a case concerning the use of juries in criminal proceedings and the right of a professional judge to bring a case before the Supreme Court if they doubt the judgement of a jury, the Austrian Constitutional Court discussed rules of criminal procedure from 1850, 1873, 1934, and their current version from 1950.[166] Reference is also made to instances where the meaning of terms can change. For instance, in the case of registering non-binary genders in Germany, the German Federal Constitutional Court held that the

[157] VfGH 26.09.2017, G 134/2017-12, G 207/2017-8 [66].
[158] ibid [52], [64].
[159] VfGH 14.07.2020, G 202/2020-20, V 408/2020-20* [100]–[106].
[160] VfGH 04.12.2017, G 258-259/2017-9 [10].
[161] ibid [11]–[12].
[162] ibid [15]–[17].
[163] BVerfG, Urteil des Ersten Senats vom 5 November 2019 - 1 BvL 7/16 - Rn (1 - 225) [128].
[164] ibid [189], [210].
[165] VfGH 27.06.2018, G 415/2017-12 [25].
[166] VfGH 27.06.2018, G 28/2018-13 [35]–[38].

usage of only 'men' and 'women' in the German Basic Law only reflected earlier societal understandings and did not limit contemporary interpretations.[167]

(ii) Evaluation: Holistic Policy Considerations

It seems that the Courts do not see the holistic consideration of policies as an interference with the separation of powers; or at least they do not see it as an interference of an unacceptable degree. However, this is not so simple. One possible criticism is that the Courts go beyond what they are asked to do in their analyses given that they might implicitly pass judgement on the policy at large. For example, when the Austrian Constitutional Court discussed the reduced times for appeals against return decisions, it did more than just assess the legislation in question but essentially judged the entire policy package as insufficient for its proclaimed goal. Furthermore, the intention behind a certain measure might not be clear, and it could be problematic for the Court to define one in its analysis and on its own accord. However, in the cases considered, the Courts usually referenced policy documents or what has been argued during the proceedings when defining a policy's aim.

The practice appears to have a positive effect on checks and balances because it, on one hand, provides a very thorough check of measures by reviewing not only the specific provision in isolation, but also by reference to its broader aims. On the other hand, this check also grants a certain leeway to the legislator by considering their actions in its entirety.

In terms of protecting fundamental rights, to take the 'holistic policy' approach might result in less protection overall, as the Courts often balance the measure at issue with others. This more relative approach has led the Courts to declare measures that present an encroachment on fundamental rights to be constitutional. Conversely, this approach also allows for the finding of violations of fundamental rights based on the wider context in which the measure is situated. Thus, measures that might be constitutional by themselves have been nevertheless declared unconstitutional by reference to legislation in similar situations or the overall policy approach (or lack thereof).

Nevertheless, the practice does not seem problematic in terms of the separation of powers, though one possible criticism is that a court might not be the most competent body to undertake policy evaluations, and therefore, decisions based on policy evaluations might be less well-received.

E. The relationship between the court and the legislator

[167] BVerfG, Beschluss des Ersten Senats vom 10 Oktober 2017 - 1 BvR 2019/16 - Rn (1 - 69) [42]–[43].

(i) Theme: The Relationship between the Court and the Legislator

This final subsection draws attention to the relationship between the Courts and the legislators. First, it considers whether the Courts, when answering political questions, side more often with the government or the applicants. Second, the cases where the government did not make any statements on the case will be explored. Ran Hirschl argues that judicial empowerment often supports political interests.[168] He says the 'source of evil' of judicialisation is the prevalence of 'self-interested, risk averse politicians'.[169] According to him, governments are only willing to allow extensive power shifts to courts if they benefit from the courts taking decisions they are unwilling to take, and which might be politically costly.[170] Another reason he discusses is political elites hoping to secure their policy preferences.[171]

Although it is outside the scope of this research to assess opinions and political preferences in individual decisions, it might be interesting to note that in the cases considered both the Austrian Constitutional Court and the German Federal Constitutional Court decided significantly more often in favour of the applicants.[172] This includes findings of (partial) unconstitutionality and constitutionally conforming interpretations which, although upholding the constitutionality of legislation, nonetheless bring about the desired change. Thus, this superficial analysis indicates that the German and Austrian Courts do not necessarily decide in favour of political elites when discussing political questions.

The assertion Hirschl makes, which is that political stakeholders defer politically salient decisions to courts, partially corresponds to the findings of this research. In three cases involving politically charged questions, the Austrian government did not make any statements during the proceedings. In all three instances, the Court either declared the legislation to be unconstitutional or provided a constitutionally conforming interpretation. The first concerned same-sex marriage.[173] The Court found the law prohibiting same-sex marriage unconstitutional. The second the possibility of registering non-binary genders, the Austrian government did not make a statement.[174] The Court did not find the legislation to be unconstitutional but provided a constitutionally conforming interpretation, which allowed for the registration of non-binary genders.[175] The third discussed the minimum pecuniary penalty for unlawful entry or stay (set at €5,000), which

[168] Ran Hirschl, '"Juristocracy"—Political, not Juridical' (2004) 13(3) The Good Society 6, 9.
[169] ibid 6.
[170] ibid 8.
[171] ibid 9.
[172] The German Federal Constitutional Court sided with the applicants 18 times and with the government 7 times. The Austrian Constitutional Court sided with the applicants 18 times (including 3 constitutionally conforming interpretations) and sided with the government 11 times.
[173] VfGH 04.12.2017, G 258-259/2017-9.
[174] VfGH 15.06.2018, G 77/2018-9 [5].
[175] VfGH 15.06.2018, G 77/2018-9 [38], [45], and [46].

was found to be unconstitutional.[176] In Germany, either the Federal government or State governments, depending on whether the case concerned Federal or State legislation, have always made statements. Only in one case discussing the possibility of officially registering non-binary genders, which was regulated by Federal law (*Personenstandsgesetz*) did the German Federal Government not make a statement (though the State Government of Thuringia did make a statement in support of the possibility of officially registering non-binary genders).[177]

(ii) Evaluation: The Relationship between the Court and the Legislator

That the Constitutional Courts tend to rule in favour of the applicants is not a problem for the principle of the separation of powers. However, as discussed above, depending on how the Courts make these decisions, they may risk overstepping their boundaries if they do not accept certain political certitudes. If the Courts show a tendency to uphold the constitutionality of the laws under review, this could indicate a lack of checks and balances. However, because the opposite is the case, the Courts clearly engage in thorough and critical scrutiny in this regard. Similarly, as to the protection of fundamental rights, the Courts indeed often find legislation to be unconstitutional because of a lack of fundamental rights protections, and therefore their practice of tending to side with the applicants indicates strong fundamental rights protection. As to prudential concerns, the Courts might be weakened and be subject to criticism if they disagree too much with the political stakeholders. However, given that the German Federal Constitutional Court has, for the past decades, enjoyed the highest levels of trust among the constitutional bodies,[178] such disagreement seems to not have affected Court negatively. Similarly, the Austrian Constitutional Court's decisions are generally accepted, and the use of constitutional review is viewed rather positively.[179]

The deferral of politically salient decisions to the Constitutional Courts, which seems to be more prevalent in Austria as indicated by the government making no statements in some proceedings, is a clear problem for the principle of the separation of powers. The Court is somewhat forced to make decisions that might be more suited for the political process, but in relation to which the government is unwilling to take a stance. That said, though such decisions might ideally be placed elsewhere, the Court is still fulfilling its obligations under the principle of checks and balances by scrutinising the laws under consideration and providing answers the government is unwilling to give. This, in the examples discussed here, is always in favour of greater human rights protection. The deferral of political questions to

[176] VfGH 10.03.2020, G 163/2019-16 ua* [37].
[177] BVerfG, Beschluss des Ersten Senats vom 10 Oktober 2017 - 1 BvR 2019/16 - Rn (1 - 69) [18].
[178] Internetredaktion der LpB BW, '70 Jahre Bundesverfassungsgericht 1951–2021' (*Landeszentrale für politische Bildung Baden-Württemberg*, 2021) <www.lpb-bw.de/bundesverfassungsgericht#c67178> accessed 12 June 2022.
[179] Lachmayer (n 83) 14.

the Courts might be viewed critically for prudential reasons as the Courts might be forced to answer questions not best placed with them. However, the Courts might also be perceived as the party that truly protects fundamental rights and makes more radical changes that have been requested from certain societal groups that politician cannot agree upon.

VI. A Political Question Doctrine for Germany and Austria

The preceding Section presented the practice of the German Federal Constitutional Court and the Austrian Constitutional Court when considering political questions based on five themes. These themes were each evaluated according to the criteria of the separation of powers, checks and balances, protection of fundamental rights, and prudential concerns. First, Section VI.A summarizes how the Courts' practices affect the selected evaluative criteria. It is clear that no practice is able to fully satisfy all criteria and that concessions will always have to be made. Nevertheless, Section VI.B explores what a political question doctrine might look like for Germany and Austria, drawing inspiration from Cohen's 'politics-reinforcing political question doctrine'.[180]

A. The courts' practices in light of the evaluative criteria

Overall, the Courts' practice yields mixed results for all evaluative criteria. Requirements of the separation of powers are often not fully met when discussing political questions. Reference to external sources may be overly restrictive of national policymaking and may amount to policymaking by the Courts, especially when they refer to non-judicial sources. That both Courts tend not to side with the government and accept the government's deferral of decisions, which is particularly the case in Austria, suggests they are quite willing to interfere in the policy arena. Constitutionally conforming interpretations and guidelines by the Courts can be problematic, though the Courts indicate that they respect the intentions of the legislature. This attitude is also reflected in the Courts' application of a legislative margin of appreciation, which supports the separation of powers. The Courts' practice of holistically considering policies may amount to political decision-making. However, it appears to be largely non-invasive into the political realm.

Court practices mostly adhere to the principle of checks and balances. The use of multiple external references enhances a thorough review of the legislation under consideration. Similarly, holistic policy considerations allow for a more in-depth, but also more nuanced, examination of the legislation. The Courts' tendency to side with the applicants further indicates thorough constitutional review.

[180] Cohen (n 23).

Deferral of questions by the government and legislature may also enhance scrutiny. The practice of providing constitutionally conforming interpretations or guidelines is an expression of scrutiny, albeit a possibly less clear one than a declaration of unconstitutionality. Finally, only the use of a legislative margin of appreciation somewhat limits checks and balances as the Courts accept certain legislative decisions and apply rather low standards of review.

Both Courts protect fundamental rights. The cases in which the Courts side with the applicants rather than the government usually involve the finding of a violation of fundamental rights. When the Courts make decisions that the government was seemingly unwilling to make, they also opt for a rights-affirming ruling. In addition, the external sources referred to by both Courts often support stronger rights protection; both Courts, the Austrian Constitutional Court in particular, extensively engage with the case law of the ECtHR. Constitutionally conforming interpretations are usually formulated in a way that secures fundamental rights. However, protection could be more comprehensive if it were enshrined in legislation. Holistic policy considerations can ensure the protection of fundamental rights because they provide a more comprehensive view of the matter. However, they can also have a limiting effect in some cases. For example, when the balancing of the various interests leads the Courts to conclude that certain rights can indeed be restricted. The use of a legislative margin of appreciation has a similar effect, where the Courts give the legislator leeway to restrict or not actively promote fundamental rights.

The Courts' practices often appear to have little regard for prudential concerns. The constitutionally conforming interpretations and guidelines offered may amount to policymaking and may be outside the purview of the Courts. Holistic policy considerations pose similar risks. However, this does not appear to impact the Courts negatively. Reference to external sources, on the other hand, pays respect to prudential concerns as it lends legitimacy to the Courts' decisions. Nevertheless, it can be viewed critically as the Courts' engagement with non-legal sources might be beyond their proper realm. Disagreeing too frequently with legislators and accepting deferred questions carries the risks of backlash, but this does not seem to be the case in either Germany or Austria. Appeal to a legislative margin of appreciation is a sign that the Courts are proceeding prudently.

B. Cohen's Politics-Reinforcing Political Questions Doctrine

Cohen's pluralist or politics-reinforcing political question doctrine aims to protect the channels of democratic debate rather than shield government decisions from judicial review.[181] He bases his doctrine on three arguments from

[181] Cohen (n 23) 32.

constitutional theory:[182] first, that a democratic constitution should guarantee that diverse voices are heard in public debate; second, that judicial review should monitor and uphold fairness and openness of the political process; and third, that courts should not shut down political debates without good reasons. According to Cohen, the US Supreme Court should refrain from exercising constitutional review when the executive branch (the President and Cabinet) and the legislative branch (Congress) are in opposition and each branch can make a credible case that it has the independent power to determine the policy in the case.[183] However, even if this is the case, the Court must assess whether judicial intervention is necessary to respond to a possible violation of an important right.[184] If intervention is necessary, the Court should intervene minimally and render a narrow decision.

C. Sketching a doctrine for Germany and Austria

It is beyond the scope of this research to propose a fleshed-out political question doctrine for Germany and Austria. I will nevertheless endeavour to offer some thoughts on what a politics-reinforcing doctrine, to borrow Cohen's term, might look like and what should be considered when drafting such a doctrine. First, it should be noted that Cohen's doctrine cannot be simply transferred to the German or Austrian context. The political branches are structured differently and supreme courts like the one in the US are somewhat different from constitutional courts like the ones in Germany and Austria. An important aspect of the political question doctrine, as discussed in the US context, is the delimitation of power between Congress and the executive. Although this question of the delimitation of power between the legislative and the executive has also been considered before the German Federal Constitutional Court and the Austrian Constitutional Court,[185] it does not play such a significant role in the Courts' decisions. Moreover, the balance of power and the functioning between the executive and the legislative branches are different. One of the reasons for this is that in Germany and Austria the executive and the legislature are 'elected' at the same time.[186] The government in both countries is usually formed by the strongest party in the first chamber of parliament (the Bundestag in Germany and the Nationalrat in Austria) and possibly one or multiple coalition partners.[187] In both countries, these parliamentary

[182] ibid 33–41.
[183] ibid 41–44, 48.
[184] ibid 44, 48.
[185] See for example BVerfG, Urteil des Zweiten Senats vom 19 September 2018 - 2 BvF 1/ 15, 2 BvF 2/15 - Rn (1 - 357) [242]–[243].
[186] Note that the president in Austria is elected separately and directly by the Austrian people (Austrian Federal Constitutional Law, art 60), as it follows a semi-presidential system. Germany has a parliamentary system where the president is elected by a Federal Convention composed of all members of the current Bundestag and an equal number of state electors (German Basic Law, art 54).
[187] Germany: arts 63(2), 64(1), 69(1); see also Bundeszentrale für politische Bildung, 'Bundesregierung' (*Bundeszentrale für politische Bildung*) <www.bpb.de/kurz-knapp/lexika/pocket-politik/16360/bundesregierung/> accessed 7 June 2022. Austria: art 70(1); see also Demokratiezentrum Wien, 'Bundes-

chambers are elected by direct popular vote.[188] Thus, political differences between the executive and the legislative are rarer, because the majority of the legislative body also forms the executive (except for the Austrian President who is elected separately). In contrast, in the US, the executive and the legislature are elected in separate elections, which can lead to greater political differences between the two.[189]

As to the differences between supreme courts and constitutional courts, it has been argued that 'it is the job of a constitutional court "to choose and impose values" as they are positioned outside of the regular court system'.[190] This perception of constitutional courts as political actors would justify them answering political questions more frequently.[191] Supreme courts are not necessarily viewed as naturally having this power. One reason why political decision-making of the US Supreme Court seems to be viewed more critically and has been limited by a political question doctrine could be its initial design as a federal court of last instance. The German and the Austrian Constitutional Courts have been designed with their task of constitutional review in mind and on the basis of a deliberate choice of the constitution drafters to create an institution to keep the legislative and executive in check.

A politics-reinforcing political question doctrine for Germany and Austria could, as a first criterion, consider whether a discussion on the matter is currently taking place in parliament. There are several influential parties in the parliaments of both countries and different lines of argument will roughly correspond to the division between government and opposition. If the matter is debated, the Courts might decide not to intervene in the political process, unless, as Cohen suggests, the matter requires intervention to protect fundamental rights. In addition, it might be useful to oblige governments to make statements in cases involving political questions that the Courts decide to answer, to avoid deferral of these questions. Furthermore, in case one of the Courts is faced with a political question, they could be given the power to initiate parliamentary discussions on the issue, or even to initiate public consultations or to propose to parliament to do so, rather than provide an answer themselves. Public consultations exist in Austria. They are non-binding votes concerning questions of fundamental importance or importance for the whole country.[192] Currently, they can only be initiated by a majority vote in the

regierung' (*Demokratiezentrum Wien*) <www.demokratiezentrum.org/bildung/angebote/lernmodule/das-politische-system/bundesregierung/> accessed 7 June 2022.

[188] Germany: Basic Law, art 38; Austria: Federal Constitutional Law, art 26.

[189] Constitution of the United States, arts 1(2), 1(3), 2(1), Amendments XII, XIV(2), XVII.

[190] David Robertson, *The Judge as Political Theorist* (Princeton University Press 2010) as cited in Tamara Ehs, 'Felix Frankfurter, Hans Kelsen, and the Practice of Judicial Review' (2013) 73 Zeitschrift für ausländisches öffentliches Recht und Völkerrecht 451, 455.

[191] Ehs (n 190) 455.

[192] Austrian Federal Constitution, art 49b.

Nationalrat and are officially issued by the Federal President.[193] A comparable mechanism in Germany does not yet exist.[194]

A political question doctrine, as outlined above, would be consistent with the principle of the separation of powers, as it aims to preserve the democratic process and leave political decisions to the legislative and executive branches or facilitate their involvement. Nevertheless, by deciding to declare a question non-justiciable, the Courts would have to engage with the relevant legislation, thereby adhering to the principle of checks and balances. Given that severe fundamental rights infringements would allow the Courts to issue a ruling on the merits despite political discussions on the matter, fundamental rights protection would be guaranteed. Finally, this approach seems to address prudential concerns, as the Courts would be less likely to overstep the mark but rather facilitate the political process.

VII. Conclusion

This article has discussed how the German Federal Constitutional Court and the Austrian Constitutional Court have addressed political questions submitted for constitutional review in their jurisprudence from 2017 to 2021. It has evaluated their respective approach in light of the separation of powers, checks and balances, fundamental rights, and concerns of judicial prudence. Generally, despite the Courts' different origins, they approach political questions very similarly. Five themes were identified to discuss the Courts' practices: (a) discussions of legislative margins of appreciation; (b) references to external sources; (c) the offering of constitutionally conforming interpretations and guidelines; (d) the application of holistic policy considerations; and (e) discussions of the relationship between the Courts, and the legislature and the executive. The evaluation of these practices yielded mixed results, with all the practices having advantages and disadvantages.

One possible way of improving the Courts' approaches to political questions is to introduce a political question doctrine. The doctrine, which originated in US constitutional law, requires a declaration of non-justiciability of certain political questions. However, there is much academic debate about the precise meaning and implications of the doctrine. Cohen has proposed a politics-reinforcing political question doctrine that aims to strengthen and facilitate the democratic process. His doctrine will require judicial restraint in situations where the democratic process is functioning well whilst still always protecting against grave human rights violations. The exact contours of a similar doctrine applied in the German or Austrian context need to be explored in further research. However, this paper has presented an outline of a possible doctrine. This includes judicial restraint for issues being debated in parliament and possibly the option for the Court to initiate political discussions or popular consultations. However, if the Courts see the risk

[193] ibid.

[194] The only somewhat comparable mechanism is a referendum concerning the revisions of the existing division into Länder under art 29 of the German Basic Law.

of serious fundamental rights infringements, they should be able to issue a decision remedying the violation, despite the initial applicability of the doctrine. If the Courts do accept the question, government should be required to make a statement in the proceedings.

The questions of how to approach increasing judicialisation and what the appropriate role of constitutional courts in an ever-changing society is are complex. Multiple perspectives can and must be considered. This research offers one novel perspective on two European courts that have served as inspiration for many courts to follow. It has hopefully provided some insight into the complexity of the matter and some ideas for how we can think about and examine constitutional courts in their multifaceted beauty.

Cambridge Law Review (2023) Vol 8, Issue 1, 104–132

Building a Bridge to a Culture of Justification: Guidelines for Designing the Standard of Proportionality in India

RUDRAKSH LAKRA*

ABSTRACT

As a standard of review, the test of proportionality is associated by its supporters with substantively strong and transparent public reason-giving, as well as with a shift to a 'culture of justification'. However, a stream of scholarship has emerged recently that explains the perceived weaknesses of the test. This has led some scholars to focus on how the standard of proportionality can be redesigned and applied in a way that addresses the concerns raised by the critics and best forwards the values associated with the test. Unfortunately, in the Indian context, where the Indian Supreme Court only recently adopted the proportionality test, there is little discussion of how the test should be designed and operated in practice. This paper therefore lays down a broad design principle that should guide this process, and then, in light of this principle, attempts to offer concrete guidance for coherently conceiving of and applying the four specific stages of the test.

I. INTRODUCTION

The proportionality test is a standard of review that courts employ to determine if the infringement of a right in question by the impugned means was justified. Originating from Germany,[1] the proportionality test has spread across the globe and gone on to become increasingly important in today's rights-based adjudication.[2] It

* Final-year law student at Jindal Global Law School. E-mail: 18jgls-rudraksh.l@jgu.edu.in. The author would like to thank Mariyam Kamil for the insightful conversations that formed the foundation for this paper. Thanks are also due to Prof Sanskriti Sanghi, Prof Sayan Mukherjee, Prof Max Steuer, Prof Balu G Nair, Prof Kai Möller, Abhijeet Shrivastava, Medha Kolanu, and Ayan Gupta.

[1] Alec Stone Sweet and Jud Mathews, 'Proportionality Balancing and Global Constitutionalism' (2008–09) 47 Columbia Journal of Transnational Law 72, 74, 97–104; Dieter Grimm, 'Proportionality in Canadian and German Constitutional Jurisprudence' (2007) 57 University of Toronto Law Journal 383, 384.

[2] Mordechai Kremnitzer, Talya Steiner, and Andrej Lang, 'Introduction' in Mordechai Kremnitzer, Talya Steiner, and Andrej Lang (eds), *Proportionality in Action Comparative and Empirical Perspectives on the Judicial Practice* (Oxford University Press 2020) 1–2; Stone Sweet and Mathews (n 1) 73–74; David Law, 'Generic

is a staple in the jurisprudence of international and supranational courts.[3] It became a central constitutional feature in the courts of Canada, South Africa, and Israel, before migrating to other countries in Europe, Asia, and Latin America.[4] The version of the test that is most commonly referred to as its four-step variant is predominantly used in Germany ('traditional proportionality test').[5] The traditional proportionality test is intended to be structured in a manner that requires the state to justify its infringing measure at each stage, and its failure to do so at any stage ends the analysis.[6] There could be several reasons to explain the popularity of this four-stage test. Its structured nature is linked to improving substantive reasoning, transparency, and public reason-giving.[7] Cohen-Eliya and Porat explain that the widespread migration of proportionality corresponds with the emerging global culture that is shifting from a culture of state authority to a 'culture of justification'.[8]

In the Indian context, the legal discourse has focused on issues with the Indian Supreme Court's ('Indian SC') application of proportionality, the arguments for adopting the traditional proportionality test in line with the reasons

Constitutional Law' (2005) 89 Minnesota Law Review 652; Kai Möller, 'Constructing the Proportionality Test: An Emerging Global Conversation' in Liora Lazarus, Christopher McCrudden, and Niels Bowels (eds), *Reasoning Rights: Comparative Judicial Engagement* (Hart Publishing 2014) 31.

[3] Stone Sweet and Mathews (n 1) 138–152; Kremnitzer, Steiner, and Lang (n 2) 1–2.

[4] For elaboration on the difference in the standard of proportionality in Germany, Canada, Israel, and South Africa see Kremnitzer, Steiner, and Lang (n 2). For the difference between the traditional proportionality test and the test applied in Colombia and Mexico see Luisa Conesa, 'The Tropicalization of Proportionality Balancing: The Colombian and Mexican Examples' (2008) Cornell Law School Inter-University Graduate Student Conference Papers 13 <https://scholarship.law.cornell.edu/lps_clacp/13> accessed 29 Aug 2022. For the difference between the traditional proportionality test and the test in Australia see Anne Carter, 'Proportionality in Australian Constitutional Law: Towards Transnationalism' (2016) 76 Heidelberg Journal of International Law 951.

[5] Kai Möller, 'Proportionality: Challenging the Critics' (2012) 10 International Journal of Constitutional Law 709, 711–716.

[6] ibid.

[7] Grimm (n 1) 397; Mattias Kumm, 'The Idea of Socratic Contestation and the Right to Justification: The Point of Rights-Based Proportionality Review' (2010) 4 Law and Ethics of Human Rights 141; Mattias Kumm, 'Constitutional Rights as Principles: On the Structure and Domain of Constitutional Justice' (2004) 2 International Journal of Constitutional Law 574, 579; David M Beatty, *The Ultimate Rule of Law* (Oxford University Press 2004) 169–200; Matthias Klatt and Moritz Meister, *The Constitutional Structure of Proportionality* (Oxford University Press 2012) 49–51, 68–70, 72–73; Kai Möller, '"Balancing as Reasoning" and the Problems of Legally Unaided Adjudication: A Rejoinder to Francisco Urbina' (2014) 12 International Journal of Constitutional Law 222; Möller, 'Proportionality: Challenging the Critics' (n 5) 727.

[8] Cohen-Eliya and others, 'Proportionality and the Culture of Justification' (2011) 59 American Journal of Comparative Law 463; Etienne Mureinik, 'A Bridge to Where? Introducing the Interim Bill of Rights' (1994) 10 South African Journal of Human Rights 31, 32. Under a culture of authority, a measure is legitimate if carried by those who have the authority to exercise power. Thus, under a culture of authority a procedural justification, such as the measure was passed by a democratically elected legislature or is linked to values set in the constitution, are sufficient. However, under a culture of justification, these justifications based on authority are only the starting point; rather, the measure must be justified based on more substantive reasons which demonstrate the rationality, necessity, and reasonableness of the measure.

above mentioned, and the benefits offered by this standard over the other standards of review, such as the reasonability test.[9] However, there has been little to no attention given to a question of equal, if not greater significance: is this four-stage design of proportionality itself replete with flaws?[10] In other words, there is an urgent need to ensure that this four-stage test is not immune from scrutiny and to begin a conversation about how the proportionality test ought to be best conceived of, and applied by, the Indian courts. One of the major criticisms of the traditional test is that the stage of 'balancing' carries the predominant weight, and the other three stages are, at most, examined in a perfunctory manner.[11] This is concerning as it limits the justificatory potential of other stages, and the stage of balancing has consistently provoked strong criticisms. This is because of the test's design and how courts examine it. This flaw, as will be discussed in-depth in Section II.B, limits the ability of the test to promote public reasoning and reduces the rigour of the test. Overall, such constraints of the test strongly undermine the proposal that this four-stage conception furthers the contemporary culture of justification.[12]

The pressing need to address the major criticisms levelled against this four-stage conception and to ensure that the test remains a valuable judicial tool necessitate this paper.[13] This paper attempts to provide meaningful insights into how the Indian SC should structure and apply the test in constitutional law. As this paper will focus on the theoretical conception of the test, it does not aim to provide a comprehensive account of how the standard of proportionality should operate in India. It will therefore not, for example, address any issues about the degree of deference and evidentiary standard that should be applied when adjudicating

[9] Chintan Chandrachud, 'Proportionality, Judicial Reasoning, and the Indian Supreme Court' (2017) 1 Anti-Discrimination Law Review 87; Abhinav Chandrachud, 'Wednesbury Reformulated: Proportionality and the Supreme Court of India' (2013) 13 Oxford University Commonwealth Law Journal 191; Vikram Aditya Narayan and Jahnavi Sindhu 'A Historical Argument for Proportionality Under the Indian Constitution' (2018) 2 Indian Law Review 51; Aparna Chandra, 'Proportionality in India: A Bridge to Nowhere?' (2020) 3 Oxford Human Rights Journal 56; Kremnitzer, Steiner, and Lang (n 2); Mariyam Kamil, 'Right to Privacy in India: Existence, Scope and Challenges' (DPhil Thesis, University of Oxford 2019) ch 6; Mariyam Kamil, 'Puttaswamy: Jury Still Out on Some Privacy Concerns?' (2017) 1 Indian Law Review 190, 197–201; Mariyam Kamil, 'The Aadhaar Judgment and the Constitution – II: On Proportionality' (*Indian Constitutional Law and Philosophy*, 30 Sept 2018) <https://indconlawphil.wordpress.com/2018/09/30/the-aadhaar-judgment-and-the-constitution-ii-on-proportionality-guest-post/> accessed 29 Aug 2022.

[10] The only work in the Indian context to my knowledge that touches upon this issue is by Duara, who argues for the adoption of the standard of proportionality for cases involving gender equality. The issue of how the test should be applied is discussed in brief. However, there is an absence of a comprehensive and systemic discussion on how the standard should be designed and applied. See Juliette G Duara, *Gender Justice and Proportionality in India Comparative Perspectives* (Routledge Advance in South Asian Studies 2018).

[11] Kremnitzer, Steiner, and Lang (n 2) 44–48, 103–11; Niels Petersen, *Proportionality and Judicial Activism Fundamental Rights Adjudication in Canada, Germany, and South Africa* (Cambridge University Press 2017) 83–98; Lazarus (n 2) 50; Grimm (n 1) 393; Stavros Tsakyrakis, 'Proportionality: An Assault on Human Rights? A Rejoinder to Madhav Khosla' (2010) 8 International Journal of Constitutional Law 307, 308; Aharon Barak, *Proportionality Constitutional Rights and their Limitations* (Cambridge University Press 2012) 339.

[12] Kremnitzer, Steiner, and Lang (n 2).

[13] Lazarus (n 2).

cases using the proportionality test, or under which exact provisions of the Constitution the proportionality test should serve as a standard of review.

In pursuing these aims, this paper also relies upon the jurisprudence of the apex courts of Germany, Canada, South Africa, and Israel. The lessons that may be learnt from these jurisdictions about the four-stage test have immense value, given that their courts have considered the test as a central adjudicatory tool.[14]

Section II explains the major faults with the theoretical design of the traditional proportionality test and its application in practice. The section first briefly explains how the traditional proportionality test is designed and the reasons offered by the advocates of this test for its support. Then, it will discuss why under the traditional test the first three stages are marginalised and the key issues with the stage of balancing carrying the predominant weight. Based on this discussion it is argued that proportionality should be designed and applied in a manner that maximises the potential of each stage, instead of marginalising any stage, to ensure that no one stage carries the predominant weight of the test.

Sections III to VI examine how adjudication should be conducted at the stages of legitimate aim, suitability, necessity, and proportionality *stricto sensu* respectively, and how these stages can be redesigned. Section III analyses two proposals for the stage of legitimate aim. The first one urges courts to require states to offer concretely defined goals; the second urges courts to examine both the objective and the subjective purpose of the state's measure. Section IV presents four proposals for the stage of suitability: first, regarding the nature of analysis the courts ought to perform at this stage; second, requiring the state to examine the measures' rational connection with the stated goals both *ex ante* and *ex post*; third, to examine the counter-productiveness of a particular measure; finally, whether the state's measure goes overboard. Section V focuses on the stage of necessity. I argue for the rejection of both the traditional proportionality test and the Blitchz standard. Thereafter, I discuss an alternative that addresses the key problems with both the standards. Section VI explains how Indian courts should adjudicate at the stage of proportionality *stricto sensu*. Moreover, it argues for the rejection of the concept of balancing adopted by the Indian SC in *Puttaswamy II*. Section VII offers concluding thoughts.

II. Examining the Drawback the Drawbacks of the Traditional Proportionality Test

The traditional proportionality test is a structured four-tier test which requires the state to sequentially meet each stage for a measure to be considered legitimate. The four stages are as follows.

[14] Kremnitzer, Steiner, and Lang (n 2); Petersen (n 11); Grimm (n 1).

The first stage is that of a 'legitimate aim'. The state's measure ought to follow a legitimate purpose.[15] This stage serves two purposes: first, it acts as a gatekeeper and weeds out state measures that follow unworthy purposes,[16] as it is critical in a constitutional democracy that rights should only be limited for constitutionally legitimate reasons.[17] Second, the test assists judges in analysing the latter stages by defining the state's goal.[18]

The second stage is 'suitability'. At this stage, we determine whether the state's measure has a rational connection with the declared worthy purpose[19]—in other words, if it can promote that goal. The point of this stage is to establish if the legitimate goal and the right in question clash.[20] If the means contribute to the achievement of the end goal, then there is a conflict between the goal and the right.[21] Conflict necessitates that one value will only be realised at the cost of the other.[22] But, if the means in question do not forward the worthy purpose at all, then there is no clash, and such a measure must fail the test.[23]

The third stage is 'necessity'. The necessity limb requires that amongst two means that can promote the state's aim to the same extent, the one that is less intrusive should be chosen.[24] This involves evaluating the effectiveness of the means in achieving the purported purpose and a relative evaluation of the degree of infringement of the rights in question by the different means.[25] The design of the necessity test is meant to ensure that any infringement of Principle 1 is only allowed strictly to the extent it is necessary to realise another vital competing Principle 2.[26] This can only be met if the least restrictive means that equally realises Principle 2 is adopted.[27] This stage helps a judge understand the scope of the policy, the level of impact of the measure on the right, and the effectiveness of the measure in achieving the State's goal. This information can provide the factual basis for conducting the balancing exercise.

The fourth stage is the balancing exercise, or what is often known as 'proportionality *stricto sensu*'. This stage determines 'whether the interference with the right is justified in light of the gain in the protection for the competing right or interest. To this end, the two values have to be 'balanced' against each other.'[28] The balancing stage is 'particularly well suited, should be tying together the analyses conducted in the previous subtests, while clearly expressing the constitutional

[15] Möller, 'Proportionality: Challenging the Critics' (n 5) 711–12; Barak (n 11) 245–46.
[16] Kremnitzer, Steiner, and Lang (n 2) 565.
[17] Kremnitzer, Steiner, and Lang (n 2) 565; Barak (n 11) 245–246.
[18] Kremnitzer, Steiner, and Lang (n 2) 565.
[19] Barak (n 11) 303.
[20] Möller, 'Proportionality: Challenging the Critics' (n 5) 713.
[21] ibid.
[22] ibid.
[23] ibid.
[24] Barak (n 11) 317–18.
[25] ibid.
[26] Lazarus (n 2) 43–44.
[27] ibid.
[28] Möller, 'Proportionality: Challenging the Critics' (n 5) 715.

values that guide the decision and the balancing considerations that lie at its foundation'.[29]

As noted above, the advocates of this test link it to a culture of justification, public reason-giving, structured reasoning, and substantive reasoning.[30] For instance, Möller explains how the test promotes structured reasoning by allowing judges to be analytical, by breaking one complex question into four relevant subquestions that can be analysed separately.[31] In a similar vein, Grimm notes that the structured nature of the test can have a disciplining and rationalising effect on the judicial decision-making process by requiring the Court to examine the state's measure stage-wise sequentially.[32] Kumm, on the other hand, credits the proportionality test with the promotion of reason-giving and transparency, arguing that proportionality is akin to the Socratic contestation method where the public authority has to justify its actions at each step by providing a public reason.[33] This proposition, however, is only attractive at first blush, because the potential of the traditional proportionality test is limited due to its design and application. This is because in the final stage proportionality *stricto sensu* carries the predominant portion of the weight and other stages are examined in a perfunctory manner.[34] This turns proportionality from a four-stage test to a balancing-centred test. This is concerning for two reasons: first, it severely limits the justificatory potential of the other stages, which can make a meaningful contribution to the test; and second, balancing as an exercise has certain theoretical issues which entail that the weight it carries should be limited in the test.[35] In this light, this section will first explain why the stages other than *stricto sensu* balancing are examined in a perfunctory manner and how this limits their justificatory potential (Section II.A), and second, it will explain the issues with the balancing stage carrying the major weight in the test (Section II.B).

[29] Kremnitzer, Steiner, and Lang (n 2) 590.

[30] Grimm (n 1) 397; Kumm (n 7); Mattias (n 7); Beatty (n 7); Klatt (n 7); Möller, 'Proportionality: Challenging the Critics' (n 5) 727; Möller, 'Balancing as Reasoning' (n 7); Cohen-Eliya (n 8).

[31] Möller, 'Proportionality: Challenging the Critics' (n 5) 727; Möller, 'Balancing as Reasoning' (n 7).

[32] Grimm (n 1) 397.

[33] Kumm (n 7).

[34] Kremnitzer, Steiner, and Lang (n 2) 44–48, 103–11; Petersen (n 11) 83–98; Lazarus (n 2) 50; Grimm (n 1) 393; Barak (n 11) 339.

[35] Concerns regarding it range from that it requires quantitatively balancing incommensurable values ('incommensurability objection'), that it results in ad hoc and impressionistic balancing and decision making ('ad hoc balancing objection'), and finally, that the stage allows the court to second-guess legislative choices ('separation of power objection'). See Francisco J Urbina, *A Critique of Proportionality and Balancing* (Cambridge University Press 2017); Francisco J Urbina, 'Is it Really That Easy? A Critique of Proportionality and "Balancing as Reasoning"' (2015) 27 Canadian Journal of Law & Jurisprudence 167; Petersen (n 11) ch 2; Stavros Tsakyrakis, 'Proportionality: An Assault on Human Rights?' (2009) 7 International Journal of Constitutional Law 468; Lazarus (n 2); Matthias Klatt and Moritz Meister, 'Proportionality—A Benefit to Human Rights? Remarks on the I·CON Controversy' (2012) 10 International Journal of Constitutional Law 687; Möller, 'Proportionality: Challenging the Critics' (n 5).

A. Marginalisation of the legitimate aim, sustainability, and necessity stages

(i) Legitimate Aim and Suitability

In mainstream literature, the first and the second stages of legitimate aim and suitability respectively are described and merely treated as threshold stages that the state can easily pass.[36] However, this does not have to be a *fait accompli* as these stages can be examined in a manner that maximises their potential by requiring the state to offer a more cogent justification. This examination can set the stage for, and even enrich the analysis at the later stages.[37] For instance, at the stage of suitability, the main inquiry is whether the measure at hand is rationally connected to the state's legitimate goal. This does not, however, tell us anything about the nature of the rational connection vis-à-vis the measure and the goal. There is a need to go beyond a mere 'means and ends' analysis to an analysis of any value-based addition offered by the state's measure.[38] We should expand the scope of our inquiry and ask, for example, to what extent the measure furthers the state's aim, whether it has a real or an illusory contribution to its purported goals, or whether the state's measure can have any parallel counter effects which would hinder its achievement of its goals.[39] By conducting inquiries like these, amongst others, the court will not only compel the state to offer more substantive reasoning for its actions at this stage, but will also assist itself in understanding the design of the state's measure and its potential to fulfil the legitimate purpose.[40] These insights will be crucial for finding suitable alternatives at the stage of necessity. Furthermore, understanding the impact and the design of the measure will help create a factual context to guide the balancing process.

Similarly, the legitimate aim stage can be examined to maximise its potential to meet the two above-mentioned purposes—to weed out unworthy goals and to set the tone for the latter stages. This can be done by requiring the state to provide concrete and well-defined goals, instead of vague or abstract goals, and by examining both the state's objective and subjective purpose.[41] These inquiries will root out any disguised unworthy goals, and having a well-defined goal will help the court in examining the other stages as well.[42] To offer an example, having a

[36] Lazarus (n 2) 91; Stavros Tsakyrakis, 'Proportionality: An Assault on Human Rights? A Rejoinder to Madhav Khosla' (2010) 8 International Journal of Constitutional Law 307, 308; Stone Sweet and Mathews (n 1) 76; Barak (n 10) 246–47, 315–17; Julian Rivers, 'Proportionality and Variable Intensity of Review' (2006) 65 Cambridge Law Journal 174, 195–98.

[37] Kremnitzer, Steiner, and Lang (n 2) 564–568, 573–576; Mordechai Kremnitzer, 'Constitutional Proportionality: (Appropriate) Guidelines?' in Gideon Sapir, Daphne Barak-Erez, and Aharon Barak (eds), *Israeli Constitutional Law in the Making* (Hart Publishing, 2013).

[38] Kremnitzer, Steiner, and Lang (n 2) 574.

[39] ibid 574–75.

[40] ibid.

[41] ibid 567.

[42] ibid.

concrete goal will help the court at the stage of suitability, where it has to determine if the impugned measure adopted by the state contributes to the realisation of the purported worthy purpose.[43] If the goal is generalised, it is more difficult to decide if the measure contributes to the goal.[44] For instance, consider a situation where the state cites national security or a threat of terrorism as a reason for a limitation without explaining the exact interest or threat that necessitated the measure. In this case, without understanding the exact goal of the state, it would be difficult to evaluate whether, and if so to what extent, the state measure contributes to the realisation of the aim. I will carry out a more in-depth analysis of how these two stages can be reformed in Sections III and IV, but it is evident from this discussion that these stages do not have to be merely threshold stages. If examined well, they can carry more weight in the test, and they can augment the quality of reasoning at the rest of the stages, thereby further promoting the culture of justification.

(ii) Necessity

Moving on to the third limb of the test, the work of Blitchz demonstrates that the stage of necessity is otiose under the traditional proportionality test because it is exceedingly difficult to find an alternative that would achieve the state's goals to the same extent.[45] This limitation, he rightly points out, significantly weakens the test.[46] In Germany, where the traditional proportionality test is followed at the necessity stage, empirical data supports Blitchz's assertion about necessity becoming otiose and highlights the dominance of the final stage of the test. In 84% of the reviewed cases, the impugned measure failed at the fourth stage, and it was only in 14% of the cases that the measure failed at the necessity stage.[47] Further, in 44% of the cases, the German Federal Constitutional Court ('German FCC') either skipped the necessity standard entirely or only glanced over it briefly.[48] Overall, the traditional proportionality test makes it difficult for the courts to find a viable alternative measure, making it virtually impossible for a measure to fail at this stage unless the court deviates from the set standard. Indeed, in certain cases where the state's measure failed at the stage of necessity, it was because the German FCC deviated from the traditional test of necessity.[49] All this shows is that the potential of the necessity stage to promote a culture of justification is rather limited because of the way it has generally been applied. Moreover, the rigour of the current test is limited, as the current design of the traditional necessity test works in a way that the most intrusive options—such as a blanket ban—would pass the test

[43] ibid.
[44] ibid.
[45] Lazarus (n 2) 42.
[46] ibid 49–51.
[47] Kremnitzer, Steiner, and Lang (n 2) 44, 580–81; Petersen (n 11) 84–98.
[48] Kremnitzer, Steiner, and Lang (n 2) 46–47.
[49] ibid 581.

because the state gets to determine the level of protection desired, and, by extension, the means necessary to achieve that level of protection. Consequently, the courts are led to endorse the most intrusive options, such as a blanket ban, because it would likely be the most effective option to achieve the level of protection desired. This issue is best highlighted by the *Adalah* decision of the Israeli Supreme Court.[50] In this case, the Israeli government had imposed a blanket ban on the unification of families where one spouse was an Israeli and the other was residing in the West Bank or the Gaza Strip. The Israeli Supreme Court found that this measure passed the necessity test because less intrusive targeted measures would not have been as effective.[51] Finally, the culture of justification is undermined as the data from Germany indicates that once it is shown that the traditional necessity stage cannot be met, examining it merely becomes a mechanical ritual for the court, entailing that it is thereby marginalised.[52] This takes the burden away from the executive and legislature to put in the effort to come up with and evaluate possible alternatives as part of their decision-making process.

B. Issues with marginalising these stages

As discussed above, because of how the test is designed and applied in practice, the last stage (balancing) carries the predominant weight in the test. This raises three concerns. First, marginalising the first three stages can impact the quality of reasoning at the stage of balancing itself, as all stages feed into the last stage, as noted above. If the other stages are examined well, the insights from those stages can help a judge gain a concrete understanding of the two competing interests in question and the relevant factors that must go into the balancing process.

Second, the balancing exercise is complex and an arduous process as it requires the balancing of potentially incommensurable constitutional values.[53] The relative worth of these values may not be capable of measurement on a set scale.[54] Assessing the competing interests of (for example) privacy and national security, the process is like comparing the length of lines to the weight of stones. Because of this issue of incommensurability, balancing cannot be carried out by quantifying

[50] HCJ 7052/03 *Adalah Legal Centre for Arab Minority Rights v Minister of Interior* [2006] IsrSC 61(2) (Israeli High Court of Justice) 2022; HCJ 8091/14 *Centre for the Defence of the Individual v Minister of Defence* (Israeli High Court of Justice, 31 December 2014, Israeli High Court of Justice), available in English at <http://versa.cardozo.yu.edu/opinions/hamoked-center-defense-individual-v-minister-defense> accessed 30 Aug 2022.

[51] *Adalah Legal Centre* (n 50) [89].

[52] Kremnitzer, Steiner, and Lang (n 2) 588.

[53] Möller, 'Proportionality: Challenging the Critics' (n 5) 719–24, 727–30; Francisco J Urbina, *A Critique of Proportionality and Balancing* (Cambridge University Press 2017) 172–75; Petersen (n 11) 40–49; Klatt (n 7) 58–64.

[54] *Bendix Autolite v Midwesco Enterprises* 486 US 888 (1988). Judge Scalia in his dictum asserted that balancing competing constitutional values is like evaluating 'whether a particular line is longer than a particular rock is heavy'.

constitutional values, as some have suggested.[55] Instead, these values can only be balanced by fashioning normative moral arguments to decide which of the values should triumph over the other in light of the circumstances of the case.[56] Even proponents of the test agree that the structure of the proportionality test provides no comprehensible guidance to aid this process.[57] This is what inspires the 'impressionistic' balancing objection, according to which under the test, no rational standards or considerations are placed that would guide how the balancing would be conducted.[58] Möller, one of the most ardent supporters of the test, agrees with this objection and explains that as a matter of moral reasoning, why a value should triumph over another can only be satisfactorily answered by creating a general and substantive moral theory of balancing that would guide this determination.[59] In the absence of a general account or a theory of rights, Möller explains that 'all of us, including judges, have no choice but to rely to some extent on our intuitions when striking a balance between a right and a competing value'.[60] Therefore, one can at the minimum conclude that the process of balancing is complex. It is inherently subjective, and it may be influenced by the value preferences of the judge and the times.[61]

In the Indian context, the Indian SC has never provided a general account of a right or substantive moral theory of balancing that would help guide the process of conducting the proportionality analysis. The judgment of the Indian SC in *Bachan Singh v State of Punjab* illustrates this point clearly. In this case, the Indian SC held that in deciding whether to impose the death penalty, a balancing exercise considering the mitigating and aggravating circumstances was to be followed.[62] However, in subsequent cases, many benches of the SC have come to different conclusions in similar circumstances, and its jurisprudence has been termed inconsistent and arbitrary.[63] This inconsistency in the process of balancing, it is argued, arises out of the difference in the value preferences of the judges adjudicating these cases.[64] In contrast to balancing, the other stages of the traditional

[55] ibid.
[56] Möller, 'Proportionality: Challenging the Critics' (n 5) 722; Urbina (n 35) 176–179.
[57] Möller, 'Balancing as Reasoning' (n 7) 223.
[58] Tsakyrakis (n 35) 482; Grégoire Webber, *The Negotiable Constitution: On the Limitation of Rights* (Cambridge University Press 2009) 89; Lazarus (n 2) 70–71.
[59] Möller, 'Proportionality: Challenging the Critics' (n 5) 728–29.
[60] ibid 728.
[61] *State of Rajasthan v Union of India* (1977) 3 SCC 592, 648: 'it is an accepted fact of constitutional interpretation that the content of justiciability changes according to how the Judge's value preferences respond to the multi-dimensional problems of the day' (Chandrachud J).
[62] *Bachan Singh v State of Punjab* AIR 1980 SC 898.
[63] Law Commission of India, 'The Death Penalty' (Report No 262, Aug 2015) <https://lawcommissionofindia.nic.in/reports/report262.pdf> accessed 29 Aug 2022; Preeti Pratishruti Dash, Neetika Vishwanath, and Anup Surendranath, 'The Enduring Gaps and Errors in Capital Sentencing in India' (*Project 39a*, September 2020) <https://www.project39a.com/op-eds/the-enduring-gaps-and-errors-in-capital-sentencing-in-india> accessed 29 Aug 2022.
[64] Law Commission of India (n 63); *Bachan Singh v State of Punjab* AIR 1980 SC 898 (Dissent J Bhagwati) [73], [76].

proportionality test are more limited and clinical in scope and are often more objective. Consequently, placing more reliance on other stages of the test will ensure that the adjudicatory process carried out by the Indian courts is more foreseeable, predictable, and consistent.

Third, in placing predominant weight on the stage of balancing, the traditional proportionality test creates an issue of separation of powers.[65] The stage of balancing allows the court to reassess and impose its considerations over the political choices made by an elected body which is supposedly better suited to understand public preferences.[66] As discussed above, judges may, while carrying out balancing, decide cases based on their value preferences and intuitions.[67] Replacing decisions made by elected leaders with a judge's personal value preferences is even more unsettling.[68] The separation of powers objection can be tempered if the balancing stage did not carry the predominant weight and if other stages, which do not raise the same concerns as balancing, carried greater weight in the test. Unlike balancing, the other stages of the test do not re-examine the balance reached by the legislature between different values. A legitimate aim and suitability are only established if the state's purpose is valid and if the measure advances the state's alleged goals. Similarly, the stage of necessity only questions whether the state could have achieved the same goal by an alternative, less intrusive means. It questions the design and the choice of the measure, and not the wisdom of the measure adopted by the legislature itself.

C. Conclusion: foundational guiding principles for redesigning the traditional proportionality test

Two insights emerge from the foregoing discussion: first, the traditional proportionality test marginalises all stages before balancing; and second, the balancing exercise raises a range of serious theoretical issues. These issues directly put into doubt the alleged potential of the traditional four-stage proportionality test to promote a culture of justification and substantive reasoning. There is, accordingly, an urgent need for the traditional proportionality test to be redesigned in a manner that would address or mitigate these concerns, given the prominence of the test in rights adjudication. In line with the work of Kremnitzer, Steiner, and Lang, the test should be designed and applied with an approach that maximises the potential of each stage instead of marginalising any stage, so to ensure that no one stage carries the predominant weight in the test.[69] Maximising the potential of

[65] Lazarus (n 2) 68–69; Tsakyrakis (n 35) 470; Klatt (n 7) 75–77; T Alexander Aleinikoff, 'Constitutional Law in the Age of Balancing' (1987) 96 Yale Law Journal 943, 984–86; Norman Siebrasse, 'The Oakes Test: An Old Ghost Impending Bold New Initiatives' (1991) 23 Ottawa Law Review 99, 107.
[66] ibid.
[67] ibid.
[68] ibid.
[69] Kremnitzer, Steiner, and Lang (n 2) 588.

each stage is important as each stage has unique contributions to make, contributions that are also critical for the analysis at the later stages.[70] That said, I do not advocate for the removal of the balancing stage, as some authors do. Rather, my only objection is with it carrying the predominant weight. The balancing process should be conducted as a last resort, to reject infringements that pass other stages of the test. The ability of the balancing process to do so is what makes the proportionality test a more rigorous standard of review than others—such as the 'strict scrutiny' standard—at times.[71] In a liberal constitutional democracy like India, the position given to fundamental rights should be paramount, and therefore the balancing exercise should be conducted despite the concerns raised. As is discussed further in Section VI, the balancing exercise can be conducted in such a manner as to address certain criticisms raised against it.

In the sections that follow, I will explain how the proportionality standard can be modelled along the lines of the broad design principles laid out above. The standard proposed below is one that would require the state to offer better justifications, and it will be stricter than the traditional proportionality test, thereby offering greater protection to fundamental rights.

III. Legitimate Aim

Under the traditional proportionality test, at this stage, the state's measure should have a legitimate purpose.[72] The Indian Constitution does not explicitly list legitimate goals under each article. For instance, under articles 14 and 21, the Constitution's text does not express legitimate goals; rather, they have been read via judicial practice, and legitimate aims have only been listed for article 19.[73]

Chandra's empirical work on the standard of review used by the Indian SC has found that the Indian SC almost always engages in the analysis of valid purpose and that this sets the stage for subsequent analysis.[74] As Chandra notes, '[t]he general balancing stage, in particular, draws heavily on the purpose inquiry, since the purpose inquiry clarifies the interests that the state is pursuing' and if a state fails to demonstrate that its measure pursues a proper aim it will probably fail.[75] Even

[70] ibid.

[71] Kamil, 'Right to Privacy in India: Existence, Scope and Challenges' (n 9); Kamil, 'The Aadhaar Judgment and the Constitution – II: On Proportionality' (n 9). ('As part of strict scrutiny, the Court answers the following two questions: 1) Is the State pursuing a 'compelling' State aim? 2) Is the State pursuing the least intrusive means of achieving its compelling objective?').

[72] Möller, 'Proportionality: Challenging the Critics' (n 5) 711–712; Barak (n 11) 245–246.

[73] Kremnitzer, Steiner, and Lang (n 2) 471.

[74] ibid 505.

[75] ibid.

recent cases that have dealt with the structured test of proportionality have included legitimate purpose as part of their test.[76]

Under extant literature, the legitimate purpose stage, as well as the suitability stage, are understood to be threshold stages that are easily passed and that have a limited contribution to make to the determination of the outcome of the proportionality analysis.[77] This is reflected even in the practice of states such as Canada, Israel, and Germany.[78] It is posited, however, that these elements of the test should not be marginalised. Rather, the unique contribution of these stages of the test (which is detailed in Section II.A) need to be maximised. If examined well, these two stages can set the tone for the rest of the analysis. For making this stage more robust and to maximise its potential, two proposals for Indian courts are provided below.

A. CONCRETELY DEFINED GOAL

Courts should ensure that the legitimate aim offered is well-defined in the sense that it is concrete and specific instead of being general, abstract, or vague.[79] In line with this, the UN Special Rapporteur Frank LaRue has noted the need for clear and precise grounds for limitations rather than 'vague and unspecified' grounds, such as broadly defined terms like 'national security' and 'terrorism'.[80] This is because broad, undefined terms might be used by the state to justify targeting vulnerable groups such as human rights defenders, journalists, or activists.[81] The concerns raised by the Special Rapporteur are exemplified in India's anti-terrorism law, the Unlawful Activities Prevention Act, 1967 (UAPA).[82] For instance, in section 15, the Act defines terrorism broadly: it does not, as best practice would dictate, limit the section to acts carried out intentionally; nor does it limit the section to certain kinds of acts. Rather, it loosely proclaims that 'acts' that threaten or are 'likely to threaten' India's unity, integrity, sovereignty, security, or economic security can be punished as terrorism.[83] Along with this, section 18 punishes not only 'inciting or conspiring the commission' of these vaguely defined terrorist acts, but goes as far as to punish people for the preparation of a

[76] Chandra (n 9); *Central Public Information Officer, Supreme Court of India v Subhash Chandra Agarwal* (2019) 8 MLJ 222 (SC) [42], [131]; *Joseph Shine v Union of India* (2019) 3 SCC 39 [279]; *Anuradha Bhasin v Union of India* (2020) 3 SCC 637 [78]–[80]; *Internet and Mobile Association of India v Reserve Bank of India* (2020) 10 SCC 274 [272]; *Akshay N Patel v Reserve Bank of India* (2022) 3 SCC 694 [32].

[77] Lazarus (n 2); Tsakyrakis (n 36); Stone Sweet and Mathews (n 1); Barak (n 10); Rivers (n 36).

[78] Kremnitzer, Steiner, and Lang (n 2) 557, 564.

[79] Kremnitzer, Steiner, and Lang (n 2) 567; Kremnitzer (n 37) 229.

[80] Frank La Rue, 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression' (Report 23rd Session, 17 April 2013) A/HRC/23/40 <https://ap.ohchr.org/documents/dpage_e.aspx?si=A/HRC/23/40> accessed 30 Aug 2022, para 58.

[81] ibid [60].

[82] Abhinav Sekhri, 'How the UAPA Repackages Ideas as Crimes' (*Article 14*, 21 April 2020) <https://www.article-14.com/post/how-the-uapa-repackages-ideas-as-crimes> accessed 29 Aug 2022.

[83] ibid.

crime. These overly broad provisions cast the net very wide: they can punish even harmless conduct.[84] For instance, the charge sheet against Sharjeel Imam, an activist, primarily relied on his possession of 'radical' literature (such as Leo Tolstoy's *War and Peace* and Christophe Jaffrelot's *Hindu Nationalism*) and his thesis as evidence for the UAPA charge.[85] Given these issues, amongst others, the UAPA has been regularly abused.[86] There is therefore a critical need to ensure that the state justifies its measures by demonstrating in a specific and individualised fashion the precise nature of the threat by offering concrete goals to limit the possibility of abuse.

Indeed, transparency will be promoted if the state offers concrete goals, as this makes it easier for courts to determine the sincerity of the state's goal; conversely, it is often easier for a state to pursue unworthy goals under the disguise of a generalised and broad valid goal.[87] Moreover, having well-defined goals can also help us understand the relative importance of that goal by a comparative analysis of other values at stake.

Further, having a concrete goal is important for the meaningful analysis of the measure at the latter stages of the traditional four-stage test. For instance, at the stage of suitability, the issue becomes whether the specific restriction adopted by the state or the means in question contribute to the realisation of the goal.[88] It is difficult to determine whether, and if so to what extent, certain means contribute towards a generalised goal, as explained in Section II.A. At the stage of necessity, the judge has to determine if the restriction over a right is more than necessary to realise the legitimate goal pursued by the state.[89] This evaluation, which is clinical in nature, is not possible until the goal at hand is specified. How can alternatives that would achieve the state's aim to the same extent be evaluated when the exact aim of the state is unclear? Likewise, at the stage of balancing (the fourth stage), the two values at hand need to be balanced in light of the concrete circumstances of the case.[90] Again, courts would likely be disabled from carrying out this balancing well if one of the values that need to be compared is left amorphous. The key point here is the concrete and detailed circumstances of the case must be considered. In the absence of such circumstances, impressionistic balancing could result. Therefore, I posit that courts ought to ensure the state's goal in question is concretely defined.

[84] ibid.
[85] Bismee Taskin, 'The Books Used as Evidence for Sedition Charges Against Sharjeel Imam, Akhil Gogoi & Others' (*The Print*, 27 September 2020) <https://theprint.in/india/the-books-used-as-evidence-for-sedition-charges-against-sharjeel-imam-akhil-gogoi-others/510728/> accessed 21 January 2013.
[86] Abhinav Sekhri, 'How the UAPA is Perverting the Idea of Justice' (*Article 14*, 16 July 2020) <https://article-14.com/post/how-the-uapa-is-perverting-india-s-justice-system> accessed 29 August 2022.
[87] Kremnitzer, Steiner, and Lang (n 2) 567; Kremnitzer (n 37) 229.
[88] Barak (n 11) 303.
[89] ibid 317–318.
[90] Möller, 'Proportionality: Challenging the Critics' (n 5) 715.

B. Objective and subjective prupose of a law

The objective purpose is the declared purpose of a law, and the subjective purpose of the law is determined by examining the motives of the lawmakers at the time of passing the law.[91] I argue that both the objective and the subjective purpose must be legitimate for a law to be valid and pass this stage.

The subjective test is important to ensure that legislators only pass a law with a proper purpose.[92] This test would help tackle a situation where the legislators have adopted a law that pursues an unworthy purpose, but they aim to disguise the law under the garb of a worthy purpose—facially content neutral laws, for example.[93] For instance, the Indian SC, in the *Anuj Garg* decision, struck down a piece of state legislation that banned women from working in any establishment in which liquor or other intoxicating substances were being consumed under the 'objective purpose' of ensuring the security of women.[94] A law such as the one in *Anuj Garg* should be struck down on the basis that the 'subjective purpose' of the law was antithetical to the idea of equality, as it victimised women by binding them to traditional cultural norms and stereotypes about morality and distinctions between the sexes under the garb of ensuring the security of women.[95]

IV. Suitability

Under the traditional proportionality test, at this stage, the means adopted by the state should have a rational connection to the state's legitimate purpose, or in other words, the means adopted should advance the purported purpose even if to a small extent.[96]

In relation to alleged violations of the right to equality, the Indian SC has a long history of applying the logic of the suitability test in the form of the 'rational nexus' standard, whereby the court has to determine whether the impugned measure indeed advances the stated purpose.[97] Even in cases involving the structured test of proportionality, suitability has constantly been part of the proposed model.[98] Chandra, in her empirical work on the standard of review adopted by the Indian SC, highlights how in almost all cases involving suitability the Indian SC decides if there is a rational nexus based on an abstract, logical connection and commonsensical reasoning rather than on concrete evidence.'[99] Accordingly, despite its

[91] Barak (n 11) 286–287.
[92] ibid 299–300.
[93] ibid.
[94] *Anuj Garg v Hotels Association of India* (2008) 3 SCC 1.
[95] ibid.
[96] Barak (n 11) 303.
[97] Kremnitzer, Steiner, and Lang (n 2) 518–519.
[98] Chandra (n 19); *Anuradha Bhasin v Union of India* (2020) 3 SCC 637 [78]–[80]; *Internet and Mobile Association of India v Reserve Bank of India* (2020) 10 SCC 274 [272]; *Akshay N Patel v Reserve Bank of India* (2022) 3 SCC 694 [32].
[99] Kremnitzer, Steiner, and Lang (n 2) 521.

strong theoretical underpinning, this stage imposes a low threshold for the state to pass in practice.

Interestingly, even with this low threshold, the government's measures often fail at this stage, thus indicating that the government does not pay sufficient attention to potential constitutional issues of particular measures.[100] Thus, in India, while the rational nexus test has contributed significantly to constitutional adjudication when the failure rate at this stage is examined; at the same time, this stage imposes a low threshold for the state, in part because the Indian SC often provides a high degree of deference to the state at this stage.[101]

The low level of the threshold and the scrutiny at this stage reflects how this stage is portrayed in the literature—namely as merely a threshold stage that is easily passed and has limited contribution to the determination of the outcome of the analysis.[102] The main purpose of the test at this stage, it is said, is to establish a relationship between means and ends. As noted in Section II.A, however, both the requirements of legitimate aim and suitability, if analysed well, can produce insights that could aid the courts in meaningfully conducting the proportionality analysis. I will now offer four ways to strengthen the suitability standard, which will also increase the possibility of a failure rate at the stage, thus furthering weight sharing between the stages.

A. Nature of analysis

The suitability stage should be used to determine whether there is a direct and not remote connection between the state's means and the concrete goals the court has identified at the first stage.[103] This stage should not be used as a mere analysis of means and ends.[104] Rather, the court should use this stage to understand the value-based addition of the state's measure to the goal.[105] A court should understand if, and to what extent, the state's measure can be effective to advance a specific and concrete goal and the factors upon which such effectiveness is dependent. The extent of the contribution of a measure should be decided on a case-to-case basis. However, this contribution should be real and not illusory.[106]

Evaluating the measure's contribution or effectiveness should allow for a more meaningful and value-based analysis. For instance, in the *Puttaswamy II* case, if the Indian SC had opted to engage with the effectiveness of the AADHAR

[100] ibid 523–524
[101] ibid 520–524
[102] Lazarus (n 2); Tsakyrakis (n 36); Stone Sweet and Mathews (n 1); Barak (n 10); Rivers (n 36).
[103] Kremnitzer, Steiner, and Lang (n 2) 754.
[104] ibid.
[105] ibid.
[106] The important question of whether the state's choice of means has to actually contribute to the goals, or whether they should have merely the potential for contribution is beyond the scope of this paper as it does not address the issue of the evidential standards regarding proportionality.

scheme in realising its alleged goals, there would have been a much more substantive discussion at the stage of suitability.[107] In that case, the Indian government had introduced the AADHAR scheme, which it alleged would aid the effective disbursement of government benefits by using a biometric authentication system, which would limit leakages from the system. The majority held that the AADHAR scheme passed the suitability analysis, and uncritically accepted the state's argument that the biometric authentication provided a unique identity that would eliminate any chance of duplication.[108] This argument, however, had been strongly contested by the petitioners, who provided evidence that under the AADHAR scheme, there was a large number of false positives and the possibility of a substantial failure rate, with the consequence that many people would not have access to benefits.[109] Had the majority in *Puttaswamy II* scrupulously examined the effectiveness of the AADHAR scheme, they would have had to engage with the evidence presented by the petitioners. After such an engagement, even if the measure did pass the stage, it would only have done so after significant examination, which would have required the state to offer more cogent reasons.

B. *EX ANTE* AND *EX POST*

Similar to observations that a law must have a legitimate aim at the time of its passing (*ex ante*) as well as throughout its lifetime (*ex post*), I postulate that the state's measure in question should have a rational connection with the specific goal in question throughout the entire time of its existence.[110]

The necessity of ensuring that the law is rationally connected to the goals *ex post* is highlighted by the case of *The Movement for Quality in Government* of the Israeli Supreme Court.[111] A mandatory draft was in effect imposed for everyone above 18 years of age in Israel. An exception was created for those students who devoted their lives to the study of the Torah (primarily the Haredi community). This law was aimed to encourage those students deferring military service to enlist in another form of national service or go into the workforce. This exception was challenged for violating the right to equality.[112] It was first examined three years

[107] Chandra (n 9) 77–78; Kamil, 'Right to Privacy in India: Existence, Scope and Challenges' (n 9); Anand Venkat, 'The Aadhaar Judgment and Reality – I: On Uniqueness' (*Indian Constitutional Law and Philosophy*, 27 September 2018) <https://indconlawphil.word press.com/2018/09/27/guest-post-the-aadhaar-judgment-and-reality-i-on-uniqueness/> accessed 30 Aug 2022; Anand Venkat, 'The Aadhaar Judgment and Reality – II: On Fallibility' (*Indian Constitutional Law and Philosophy*, 30 September 2018) <https://indconlawphil.wordpress.com/2018/09/30/the-aadhaar-judgment-and-reality-ii-on-fallibility/> accessed 30 Aug 2022.

[108] ibid.

[109] ibid.

[110] Barak (n 11) 312–315; Kremnitzer, Steiner, and Lang (n 2) 754.

[111] HCJ 6427/02 *The Movement for Quality Government v Knesset* [2006] IsrSC 61(1) 610 (Israeli High Court of Justice); also see David Ellenson, 'The Supreme Court, Yeshiva Students, and Military Conscription: Judicial Review, the Grunis Dissent, and its Implications for Israeli Democracy and Law' (2018) 23 Israel Studies 197.

[112] ibid.

after the legislature had passed it, and the Israeli Supreme Court held that the data highlighted the lack of rational connection between the means and the purported goal—to increase Haredi participation in the national service or the workforce.[113] The Court categorically held that when a measure is reviewed *ex post* such an 'examination should be done, in this context, not as a theoretical exercise but as a practical matter, tested by its actual results'.[114] Thus, an *ex post* review should be rooted in evidence about the effectiveness of the means.[115] The policy was not immediately struck down as the Israeli Supreme Court accepted the state's argument that a broader frame of reference was required for the effectiveness of the measure. Nevertheless, the second round of review was conducted 10 years after the measure was introduced, and this time the data again pointed to the lack of rational connection as there was no major change at the ground level and thus the exemption was struck down.[116] All this shows that if an *ex post* review is not conducted, even those measures that are ineffective or have no rational connection to the goal at hand may be treated as legitimate.

C. Overinclusive means

Third, at this stage, state measures that go overboard should be struck down. 'Overboard' here refers to those measures where it is possible to differentiate between the parts of the measure that further the legitimate goal and the parts of the measure that are not rationally connected to the goal and consequently, will not contribute to the achievement of the goal ('over-inclusive means'). In such circumstances, the parts of the measure that have no rational connection should be struck down. The rationale for this is that at this stage the means and the goal should be in conflict (so that further evaluation at the balancing stage is necessary), and when there is no connection, there is no conflict.[117]

An example of this approach is seen in the decision of the Court of Justice of the European Union in the *Digital Rights Ireland* case, where it struck down a directive that allowed for the blanket retention of data for the legitimate aim of combating 'serious crime'.[118] The Court struck down the overinclusive measure, noting that the blanket ban was 'an interference with the fundamental rights of practically the entire European population, including 'persons for whom there is no evidence capable of suggesting that their conduct might have a link, even an indirect or remote one, with serious crime'.[119]

[113] ibid.
[114] ibid 63–64.
[115] The question of whether the measure should have real or likelihood of effectiveness both at the stages of *ex post facto* review and *ex ante* review is beyond the scope of this paper.
[116] HCJ 1877/14 *The Movement for Quality Government v Knesset* (Israeli High Court of Justice, 12 September 2017).
[117] Möller, 'Proportionality: Challenging the Critics' (n 5) 713.
[118] Joined Cases C-293/12 and C-594/12, *Digital Rights Ireland Ltd v Minister for Communications, Marine and Natural Resources* EU:C:2014:238, [2015] QB 127.
[119] ibid paras 29–39.

An objection may be raised against this proposal, in that the issue of overinclusive means should be left to the necessity stage.[120] Of course, overinclusive measures can be struck down at the necessity stage. However, examining overinclusive means at the stage of suitability can provide this stage with more bite and can potentially increase the failure rate at the stage. It gives the suitability stage meaningful weight and incentivises courts to examine it properly. Conversely, when a single element is dominant in the proportionality analysis, courts tend to focus on that stage and conduct the analysis of other stages cursorily. Finally, for a court to decide if a measure is overboard, they would need to examine the design of the measure and the extent to which it meets the measure's goal, and this is evidently best done at the suitability stage. These insights are also valuable information that would help in the latter stages, as explained above.

D. Counterproductiveness

Fourth, courts should evaluate not only the effectiveness of the impugned measure, but also consider its counterproductiveness.[121] This requires judges to examine whether the rights-infringing measure has a parallel effect that could hinder the achievement of its declared goal.[122] A rational connection cannot merely concern the effectiveness of the state's measure to advance the declared goal. It should, rather, also try to determine if the measure is suitable in the sense that it would not have a counterproductive impact on the legitimate goal in question. The need is to shift from a means and ends analysis to a value-addition-based analysis.

Two examples illustrate this point: the S*tate of Maharashtra v Indian Hotels and Restaurants Association* decision of the Indian SC and the *Centre for the Defence of the Individual* decision of the Israeli Supreme Court. In both these cases, the state's measure had a counterproductive impact on the state's purported aim, and was therefore not rationally connected to the worthy purpose in question.

In *State of Maharashtra v Indian Hotels and Restaurants Association*, the Indian SC struck down a state law that barred dance performances in bars and restaurants.[123] The law was aimed at protecting women from exploitative and derogatory practices in the entertainment industry. The Court emphasised that the law had led to the unemployment of around 75,000 women, many of whom had to undertake sex work to sustain themselves.[124] In *Centre for the Defence of the Individual*, the Israeli Supreme Court highlighted that a decision to destroy the houses of innocent family members of terrorists, rather than having a deterrent effect, would

[120] Möller, 'Proportionality: Challenging the Critics' (n 5) 713; Barak (n 11) 335–337.
[121] Kremnitzer (n 37) 233.
[122] ibid.
[123] *State of Maharashtra v Indian Hotels and Restaurants Association* (2013) 8 SCC 519.
[124] ibid.

instead reaffirm any motivations to carry out acts of terror, thereby frustrating the very purpose of the measure.[125]

An objection may be raised against the introduction of the element of counterproductiveness at the stage of suitability, in that doing so could introduce balancing at the stage. The proponents of the traditional proportionality test may argue that balancing should be left for the last stage, and that an assessment of counterproductiveness would conflate the second and fourth stages, and indeed lead to the expansion of the proportionality *stricto sensu* stage (which I argued against in Section II above). It is, however, important to note that even though some form of balancing is being introduced at the stage of suitability under this proposal, the nature of this balancing is qualitatively different from the balancing carried out at the stage of proportionality *stricto sensu* and the difference means that they do not pose the same issues. Importantly, an evaluation of the trade-off between a value's positive and negative impact does not warrant the same criticism as the balancing of two competing incommensurable values. This is because the positive and negative impacts of a measure on a particular value can be assessed by reference to a single scale—that is, whether or not the value is enhanced. For instance, a conception of the right to privacy is itself sufficient for us to balance the potential positive and negative impact of a measure on the right in particular circumstances.[126] Therefore, the objection is unmerited and in fact, making the stage of suitability stricter would ensure that it carries more weight, in turn limiting the role of the proportionality *stricto sensu* stage.

V. Necessity

Under the traditional proportionality test, the necessity limb requires that the least intrusive option should be chosen amongst those that fulfil the state's objective to the same extent.[127] The Indian SC has not adopted a consistent understanding of the necessity element in its case law. It rarely applies the necessity element, and it seldom even examines if there are alternatives to the state's means because it is unwilling to second guess executive or/and legislative choices.[128] In cases where the necessity inquiry was carried out, different approaches have been taken in different judgments and by different judges. For instance, the majority in the *Puttaswamy II* case adopted Blitchz's understanding of necessity ('Blitchz standard'),[129] which is detailed below in Section V.A. On the other hand, in his dissent Justice Chandrachud adopted the necessity test as expressed in the traditional four-stage proportionality test.[130] In *Anuradha Bhasin v Union of India*, *Internet and*

[125] HCJ 8091/14 *Centre for the Defence of the Individual v Minister of Defence* (Israeli High Court of Justice, 31 December 2014).
[126] David E Pozen, 'Privacy-Privacy Tradeoffs' (2015) 83 University of Chicago Law Review 221.
[127] Barak (n 11) 317–18.
[128] Kremnitzer, Steiner, and Lang (n 2) 529–530.
[129] Kamil, 'Right to Privacy in India: Existence, Scope and Challenges' (n 9).
[130] *Justice KS Puttaswamy v Union of India* (2019) 1 SCC 1 (Chandrachud J) [653]ff.

Mobile Association of India v Reserve Bank of India and *Akshay N Patel v Reserve Bank of India*, the Court examined if there was an alternative, but it did not clarify to what extent an alternative had to achieve the state's aim.[131] Finally, in Justice Chandrachud's opinion in *Puttaswamy I*, which has been cited and followed in some cases, the necessity query was absent altogether.[132]

Thus, the primary need is for the Indian SC to adopt a clear and consistent understanding of this limb of the test. The standard that the courts adopt should be in line with the special value accorded to fundamental rights and should promote public reason-giving.

It is not suitable for the Indian courts to adopt the standard of necessity under the traditional proportionality test, for the reasons discussed in Section II.B. As argued above, the test of necessity is otiose under the traditional proportionality test as it is often very difficult to find alternatives that meet that state's aim to the same extent. This impacts the potential of the stage to promote reasons giving and reduces the rigour of the test. To address this issue raised by the necessity under the traditional proportionality test, Blitchz offers an alternative standard of necessity. This section first will explain why this alternative standard is also not the appropriate test for the Indian courts to apply. It then offers a third alternative that fixes the issues of the necessity stage under the traditional proportionality test and the Blitchz standard of necessity.

A. THE BLITCHZ STANDARD OF NECESSITY

According to Blitchz, there are four components (or limbs) of necessity under the traditional proportionality test, with this being a conception that should be reformed so to address the issues with the necessity stage.[133] The first limb is 'possibility', according to which the entire range of possible alternative measures which could achieve the government's objective had to be identified.[134] For Blitchz, this was not suitable at the stage of necessity as a wide range of measures, which may not even be practical, could be imagined.[135] Blitchz instead argues that at this stage, only those alternatives that are practically feasible need to be identified as a choice against the government's impugned measure.[136] The second limb is the instrumentality of identified alternatives, in that only those which are 'equally effective' in realising the state's objective should be retained for the purposes of comparison.[137] As explained in Section II.A , according to Blitchz, at the stage of

[131] *Anuradha Bhasin v Union of India* (2020) 3 SCC 637 [78]–[80]; *Internet and Mobile Association of India v Reserve Bank of India* (2020) 10 SCC 274 [272]; *Akshay N Patel v Reserve Bank of India* (2022) 3 SCC 694 [50].
[132] Chandra (n 9) 70–72; *Central Public Information Officer, Supreme Court of India v Subhash Chandra Agarwal* (2019) 8 MLJ 222 (SC) [42], [131]; *Joseph Shine v Union of India* (2019) 3 SCC 39 [279].
[133] Lazarus (n 2); Kamil, 'Right to Privacy in India: Existence, Scope and Challenges' (n 9).
[134] Lazarus (n 2) 51–53; Kamil, 'Right to Privacy in India: Existence, Scope and Challenges' (n 9).
[135] ibid.
[136] ibid.
[137] Lazarus (n 2) 51, 53–55; Kamil, 'Right to Privacy in India: Existence, Scope and Challenges' (n 9).

necessity, it is difficult for potential alternatives to be identified as it is difficult to find alternatives that meet the state's legitimate aim to the same extent. Blitchz notes that this problem can make the entire inquiry meaningless.[138] As an alternative, Blitchz proposes that at this stage concerning identified alternatives, only those alternatives that realise the government's aim in a 'real and substantial' manner are to be retained.[139] The third limb is impact.[140] Here, 'the differing impact upon fundamental rights of the measure and the alternatives identified' must be examined.[141] The final limb is the comparative component.[142] At this limb, building upon the findings of the second and the third limbs, the least restrictive measure that achieves the state's aim equally effectively should be selected. Blitchz argues that instead, at this stage the need is to select the 'best possible' alternative and this decision is to be made considering two factors: how the alternative realises the objective and its impact on fundamental rights.[143]

The major problem with the standard advanced by Blitchz is the design of the last limb of his model. At the last limb, the Blitchz model introduces balancing (which Blitchz agrees with) at the stage of necessity, yet doing so raises important theoretical issues as discussed in Section II.B.[144] The nature of the balancing that the Blitchz standard introduces is much closer to the stage of proportionality *stricto sensu* as it requires the degree of achievement and degree of impact to be balanced (these being the two factors one balances at proportionality *stricto sensu*).[145] This then raises an issue about the role of the stage of proportionality *stricto sensu*: when balancing is being carried out at the stage of necessity, then what is the scope and role of proportionality *stricto sensu*? The Blitchz standard, therefore, conflates the third and fourth stages of the traditional proportionality test.[146] Further, the standard is in direct conflict with the fourth stage of the test laid down by the Indian SC in the *Puttaswamy II* judgement, which is discussed further in the next section.[147]

Instead, I postulate that at the stage of necessity, the formulation provided by the Canadian Supreme Court in *Alberta v Hutterian Brethren of Wilson Company* should be followed. According to this approach, at the stage of necessity, any less restrictive alternative that achieves the state's aim to a 'real and substantial degree'

[138] ibid.
[139] ibid.
[140] Lazarus (n 2) 51, 55–56; Kamil, 'Right to Privacy in India: Existence, Scope and Challenges' (n 9).
[141] ibid.
[142] Lazarus (n 2) 51, 56–57; Kamil, 'Right to Privacy in India: Existence, Scope and Challenges' (n 9).
[143] ibid.
[144] Kamil, 'Right to Privacy in India: Existence, Scope and Challenges' (n 9).
[145] Lazarus (n 2) 56; Kamil, 'Right to Privacy in India: Existence, Scope and Challenges' (n 9).
[146] Kamil, 'Right to Privacy in India: Existence, Scope and Challenges' (n 9).
[147] ibid. For an alternative opinion see Lazarus (n 2); Petersen (n 11) ch 2.

should be adopted ('*Hutterian* model of necessity').[148] The *Hutterian* model of necessity would resolve the primary issues highlighted in the traditional proportionality test. Allowing those alternatives that achieve the state's aim to a substantial extent to be considered reduces the strictness of the necessity stage, and it would help operationalise the stage by making potential alternatives available for a court to examine.

Further, the *Hutterian* model of necessity offers two advantages over the Blitchz standard: it keeps the necessity and balancing stages separate and does not conflate them, and makes the test more rigorous. This is because the Blitchz standard does not require the least restrictive means that would achieve the state's objective in a real and substantial manner to be adopted in every case. Remodelling the necessity test in this way resolves the issue of the test being too weak by allowing real and meaningful alternatives to be considered. There are also other benefits of the necessity stage being robust, such as that meaningful engagement at this stage will provide insights that are critical for the balancing stage even if the measure passes the necessity stage. A robust necessity stage would allow the analytical burden of the proportionality test to be shared more equally between different stages and limit the role of the balancing stage. However, importantly, it does not completely devour the stage of balancing. Even if the state measure was the least restrictive and therefore compliant with necessity, it can still fail at the stage of balancing if the result is morally unjustifiable within the state's constitutional scheme. For instance, even if torture is the only way for the state to foil a terrorist plan and therefore necessary for national security and the prevention of terrorism, it would nevertheless be impermissible—because as the Indian SC held in *Francis Coralie Mullin v Union Administrator*, torture or cruel, inhuman, or degrading treatment can never be reasonable under articles 14 and 21 of the Indian Constitution.

VI. PROPORTIONALITY *STRICTO SENSU* AND BALANCING

Finally, we come to the stage of balancing. At this stage, the question becomes 'whether the interference with the right is justified considering the gain in the protection for the competing right or interest'. To this end, the two values have to be 'balanced' against each other.[149] This stage allows for those disproportionate

[148] *Alberta v Hutterian Brethren of Wilson Colony* [2009] 2 SCR 567 [55]. An objection can be raised against the adoption of the standard laid down by the Canadian Supreme Court as in Canada the literature suggests that the necessity stage carries the predominant weight and the other stages are marginalised. This could make necessity (instead of balancing) the dominant stage of the test which devours the other stages. This would lead to similar issues of a single element becoming the dominant one, as discussed above. However, none of the literature which argues that the Canadian test makes the necessity stage the predominant stage under the test applies to the *Hutterian* model of necessity, as this formation was first laid down by the Canadian Supreme Court in 2009. This formulation has only been used in five cases after that. Therefore, it is not the reason why the necessity stage devours the fourth stage in the Canadian jurisdiction.

[149] Möller, 'Proportionality: Challenging the Critics' (n 5) 715.

infringements that pass through the earlier stages to be stuck down. It is because of this that at times proportionality is, at times, a more rigorous standard of review than others, such as strict scrutiny.[150]

Thinking of the four stages as part of the same inquiry helps us understand how the balancing is enriched by the analysis of the previous stages (if done well, as I have argued in the previous sections). The earlier stages would help the court gain a concrete understanding of the two competing interests in question, the importance and the sincerity of the state's goal, the design of the measure, the actual contribution and the extent of the impact of the state's measure, and the possible alternatives and their effectiveness. These insights would then guide the judges by offering a cogent understanding of the relevant factors that should be articulated in the process of balancing.

The jurisprudence of the Indian SC has been inconsistent regarding the place of proportionality *stricto sensu*. In *Modern Dental College* and *Internet and Mobile Association*, the Indian SC adopted proportionality *stricto sensu* as the last element of the test;[151] in the *Puttaswamy II* case, the Indian SC adopted the model laid down by von Bernstorff for the last stage, (more on this in Section VI.B);[152] and in most other cases either proportionality *stricto sensu* was absent[153] or it was unclear if it was part of the test.[154] As with the stage of necessity, there is an urgent need for the Indian SC to lay down a clear standard for this stage.

This section is divided into two parts. The first explains how the analysis should be carried out at the stage of balancing and the second flags concerns with the fourth stage of the test as understood by the Indian SC in the *Puttaswamy II* case.

A. Nature of analysis

There are two broad forms of balancing that courts can conduct: the first is 'interest balancing' and the second is 'balancing as reasoning'.[155] Interest balancing is useful when the objects to be compared exist on the same scale. Interest balancing can help us decide whether we should buy apples for shop A, which sells them for 10 INR, or shop B, which sells them for 15 INR. However, it cannot help us decide whether we should buy apples or oranges if both have the same price. This is because their two objects are incommensurable. They do not exist on a common

[150] Kamil, 'The Aadhaar Judgment and the Constitution – II: On Proportionality' (n 9).
[151] *Modern Dental College and Research Centre v State of Madhya Pradesh* (2016) 7 SCC 353 [65]; *Internet and Mobile Association of India v Reserve Bank of India* (2020) 10 SCC 274 [272].
[152] Kamil, 'Right to Privacy in India: Existence, Scope and Challenges' (n 9).
[153] *Anuradha Bhasin v Union of India* (2020) 3 SCC 637 [78]–[80].
[154] Kamil, 'Puttaswamy: Jury Still Out on Some Privacy Concerns?' (n 9) 197–201 (There was a lack of clarity over the element of balancing in J Chandrachud's plurality opinion in *Puttaswamy I* which has been adopted in Joseph Shine and Subhash Chandra Agarwal decision by the Indian SC; *Justice KS Puttaswamy v Union of India* (2019) 1 SCC 1 (Chandrachud J) [653]ff.
[155] Möller, 'Proportionality: Challenging the Critics' (n 5) 715–716; Kai Möller, *The Global Model of Constitutional Rights* (Oxford University Press 2012) ch 6.

scale. As mentioned earlier, constitutional values and interests are incommensurable, and thus interest balancing cannot be applied.[156] This issue can be resolved by creating a relation between the two incommensurable objects to compare them. One might decide whether to buy oranges or apples with the same price based on the utility they will derive out of them. Similarly, to compare incommensurable constitutional values, 'balancing as reasoning'—which requires us to 'make a moral argument as to which of the competing interests takes priority in the case at hand' taking all relevant moral and practical considerations into account—is required.[157] Similarly, Kumm advocates for open-ended practical reasoning at the stage of balancing.[158] Kumm further explains that this form of reasoning requires us to assess 'whether a public action can be demonstrably justified by reasons that are appropriate in a liberal democracy'.[159] In other words, at this stage judges need to decide cases by creating a moral argument to decide which of the incommensurable values should be favoured in light of the circumstances of the case.

However, as discussed above, it is difficult to answer whether a value should be valued over another satisfactorily in the absence of a general account of right or a theory of right.[160] Examples of such general accounts lie in Möller's work, which argues that dignitarian principles such as intrinsic value, moral autonomy, and fundamental equality form the foundation for most conditional rights protection.[161] Accordingly, at the stage of balancing, when two constitutional interests need to be balanced, a theoretical foundation (that aligns with India's constitutional framework) of the interests in question ought to be developed. A coherent and meaningful account of interests would help us understand the values at stake and the values that are central to our constitutional regime. Such a framework would assist the Indian SC in the process of balancing.

B. *PUTTASWAMY II*: THE HYBRID MODEL OF PROPORTIONALITY

The Indian SC in *Puttaswamy II* adopted what Kamil calls the 'hybrid' model of proportionality—that is, the traditional proportionality test with the Blitchz standard at the stage of necessity and von Bernstorff's proposal at the last stage.[162]

[156] Möller, 'Proportionality: Challenging the Critics' (n 5) 719–724; Niels Petersen, 'How to Compare the Length of Lines to the Weight of Stones: Balancing and the Resolution of Value Conflicts in Constitutional Law' (2019) 14 German Law Journal 1387, 1389–1392; Urbina (n 35) 172–175.

[157] Möller, 'Proportionality: Challenging the Critics' (n 5) 715–16; Urbina (n 35) 176–179.

[158] Kumm (n 7) 146–47, 150.

[159] ibid 150.

[160] Möller, 'Proportionality: Challenging the Critics' (n 5) 728–29; In the sphere of public international law in the context of admissibility of evidence see Abhijeet Shrivastava, 'Two Shades of Impunity? The Balancing of Sovereign Interests in Admitting Illegally Obtained Evidence' (2023) 6 De Lege Ferenda (forthcoming).

[161] Kai Möller, 'Beyond Reasonableness: The Dignitarian Structure of Human and Constitutional Rights' (2021) 34 Canadian Journal of Law & Jurisprudence 341; Möller, *The Global Model of Constitutional Rights* (n 155).

[162] Kamil, 'Right to Privacy in India: Existence, Scope and Challenges' (n 9); *Puttaswamy v Union of India* (2019) 1 SCC 1.

This model of proportionality has not yet been adopted by any of the succeeding cases dealing with proportionality.[163] However, if this model is adopted, certain concerns with it need to be addressed.[164] The Indian SC would need to provide better justifications about why this model is appropriate and what benefits it offers over proportionality *stricto sensu*.[165]

First, as Kamil explains, there is an inherent contradiction in the work of Blitchz and von Bernstorff that makes the hybrid model of proportionality inoperable.[166] As discussed in Section V.A, the Blitchz model introduces balancing that is akin (to an extent) to the balancing process at the stage of proportionality *stricto sensu*, as it requires the degree of achievement and degree of impact of different alternatives to be balanced.[167] On the other hand, for von Bernstorff the main issue with the traditional proportionality test is with the stage of balancing which he argues should be rejected altogether.[168] Von Bernstorff argues that only the first three stages of the traditional proportionality test should be used in most cases.[169] In certain cases, such as when there is a serious infringement of rights, the fourth stage is to be replaced with categorical rules (bright lines).[170] Therefore, balancing is to be replaced with categorical reasoning. Categorical reasoning is premised on the idea that rights are rules instead of principles. This requires the creation of standards or bright lines which determine how a specific situation is decided in every case, instead of carrying out balancing every time.[171] An example of this would be if the court decided that a property right cannot trump the right to life. Thus, life could never be deprived to protect private property by law enforcement irrespective of the degree of the threat to private property. For von Bernstorff, these bright-line rules are to be constructed by the Court by a 'reference to the "essence", "substance" or "core" of a particular right *ex negativo* for specific groups of case scenarios, or by other generalisable tests or "intervention thresholds", such as the famous Brandenburg test of the US Supreme Court'.[172] Consequently, the work of Blitchz and von Bernstorff are contradictory.[173] The Indian SC's uncritical acceptance of the hybrid model of proportionality is therefore problematic.[174]

Second, apart from the inherent contradiction at the third and fourth stages of the hybrid model of proportionality, there is a lack of clarity about how the balancing stage is to be applied. This is because, with respect, the Indian SC has adopted von Bernstorff's work without in fact understanding von Bernstorff's

[163] Kamil, 'Right to Privacy in India: Existence, Scope and Challenges' (n 9).
[164] ibid.
[165] ibid.
[166] ibid.
[167] ibid.
[168] Lazarus (n 2); Kamil, 'Right to Privacy in India: Existence, Scope and Challenges' (n 9).
[169] ibid.
[170] ibid.
[171] ibid.
[172] ibid.
[173] Kamil, 'Right to Privacy in India: Existence, Scope and Challenges' (n 9).
[174] ibid.

work.[175] In *Puttaswamy II*, the Indian SC held that the process of balancing was to be conducted following bright-line rules which were either to be established or needed to be evolved.[176] The establishment of such bright-line rules was to guide the process of balancing to ensure that balancing is not conducted impressionistically.[177] Yet, von Bernstorff does not conceptualise bright-line rules as guidelines; rather, they are supposed to promote categorical reasoning.[178] This misunderstanding of von Bernstorff's work further raises issues about how the last stage is to operate.[179] Moreover, the majority in *Puttaswamy II* did not explain what the proposed bright-line rules were or how they were to be created, adding to the confusion.[180] Perhaps it is because of this issue that the *Puttaswamy II* standard has not been used in any subsequent case.[181]

Third, it is unclear if adopting von Bernstorff's work at the last stage would offer any advantage over proportionality *stricto sensu*. Von Bernstorff's main reason for promoting categorical reasoning is that it reduces the possibility of uncertainty, whereas the traditional proportionality test promotes ad hoc reasoning.[182] The courts need to consider whether legal certainty is an ontological value worth pursuing at the cost of adjudication that attempts to provide just and fair results by taking concrete circumstances into account.[183] Categorical rules are necessarily inflexible, and can therefore become over- and under-inclusive over time or produce sub-optimal results when they are applied outside the specific context in which they were created.[184] The courts would need to decide if this trade-off is acceptable.

In this regard, predictability is an important value, but it cannot devour the fairness of adjudication that should be central in a liberal constitutional democracy. The Constitution ought to be treated as a living instrument, and thus cases should be decided in the light of present-day circumstances. This might, at first blush, reduce predictability. But a level of uncertainty is arguably a warranted trade-off in return for flexible and just decision-making on the basis of unique factors and circumstances. It might also be said that the consideration of present-day circumstances also leads to a degree of certainty.

[175] ibid.
[176] Kamil, 'Right to Privacy in India: Existence, Scope and Challenges' (n 9); *Puttaswamy v Union of India* (2019) 1 SCC 1.
[177] ibid.
[178] Lazarus (n 2) 67, 83–84.
[179] Kamil, 'Right to Privacy in India: Existence, Scope and Challenges' (n 9).
[180] ibid.
[181] ibid.
[182] Lazarus (n 2) 67, 83–84.
[183] Klatt (n 7) 49–51; Petersen (n 11) 54–58; Frederick Schauer, *Playing by the Rules: A Philosophical Examination of Rule-Based Decision-Making in Law and Life* (Clarendon Press 1993) 140.
[184] Schauer (n 183).

Further, it must be emphasised that there is only a level of uncertainty; it is not the case that each case is decided in an ad hoc manner, as critics suggest. Balancing does not exclude reliance on precedents.[185] Under balancing, it is only when factors and circumstances differ would a precedent be distinguished. For instance, the South African Constitutional Court had to decide on the issue of the constitutionality of reverse onus clauses in several cases. Only in the first case, *S v Mbatha,* did the Court rely on the proportionality test; in subsequent cases, they merely cited *S v Mbatha*.[186]

Moreover, removing the last stage from the traditional proportionality test will, in most cases, reduce its rigour. As argued above, the last stage can be used to reject measures that pass the other stages of the test. Von Bernstorff's proposal to remove the proportionality *stricto sensu* stage—which, as discussed, carries the most weight under the traditional model—from the proportionality test would significantly lower the strictness of the test. In a liberal democracy where rights are fundamental, rights should be protected as far as possible.

VII. Conclusion

This paper has attempted to contribute to the emerging stream of literature, the basic assumption of which is that 'proportionality is a valuable doctrine' and should be viewed in the best light by focusing 'on the way it operates or ought to operate in practice, in the actual resolution of cases'.[187] This article has attempted to elaborate on how the traditional understanding of proportionality can be redesigned and applied in a manner that best promotes a 'culture of justification'.[188] I have argued that the test should be designed, and adjudication should take place, in a manner that maximises the justification potential of each stage and limits the role played by the last stage—that is, proportionality *stricto sensu* or balancing.[189] Based on this broad architectural principle, this article has provided certain recommendations to Indian courts regarding how each stage of the test can be redesigned. As the Indian courts have only started applying the proportionality test relatively recently and do not have the same experience as other jurisdictions, this guidance will hopefully prove helpful.

This work and the broader discourse is especially important in the Indian context as the Indian SC has failed to lay down a coherent standard of proportionality and has repeatedly failed to apply the standard scrupulously.[190] The Indian SC has so far followed a 'business as usual' approach regarding proportionality,

[185] Klatt (n 7) 49–51; Petersen (n 11) 56–57.
[186] *S v Mbatha, S v Prinsloo* (CCT19/95, CCT35/95) [1996] ZACC 1; *S v Ntsele* (CCT25/97) [1997] ZACC 14; *S v Mello* (CCT5/98) [1998] ZACC 7.
[187] Lazarus (n 2) 32–33.
[188] ibid.
[189] Kremnitzer, Steiner, and Lang (n 2).
[190] ibid.

and has refused to unsettle or disrupt 'pre-existing configurations of relations between citizens and the State as mediated through rights' through the adoption of the proportionality test.[191]

The current Chief Justice of India, in his now famous opinion in the 2018 case of *Puttaswamy v Union of India*, declared that proportionality reflected a bridge from a culture of authority to a culture of justification.[192] Unfortunately, this declaration is still far from being materialised in concrete cases where the liberty of citizens is at stake.[193] To ensure that proportionality does not turn into a 'bridge to nowhere',[194] it is imperative that the courts lay down a cogent standard of proportionality which best captures the ethos of the culture of justification.

[191] Chandra (n 9) 86; Rudraksh Lakra, 'Melancholy Takeaways on Proportionality from the Demonetisation Case' (*Indian Constitutional Law and Philosophy*, 26 Jan 2023) <https://indconlawphil.wordpress.com/2023/01/26/guest-post-business-as-usual-melancholy-takeaways-on-proportionality-from-the-demonetisation-case/> accessed 4 February 2023.

[192] *Justice KS Puttaswamy v Union of India* (2019) 1 SCC 1, 814 (Chandrachud J).

[193] Chandra (n 9); Kamil, 'Right to Privacy in India: Existence, Scope and Challenges' (n 9); Kamil, 'The Aadhaar Judgment and the Constitution – II: On Proportionality' (n 9).

[194] Chandra (n 9) 86.

Cambridge Law Review (2023) Vol 8, Issue 1, 133–151

Strengthening Women's Right to Property Acquired During Marriage: A Study of Ghana's Legal Framework

PRISCILLA AKUA VITOH*

ABSTRACT

Over the past ten years, inclusive capitalist discourses have promoted land privatisation and individual land ownership as a means of empowering women and fostering economic prosperity. This paper draws on feminist arguments that the law, as experienced by women, cannot be adequately explained by focusing exclusively on male-centric state law. It situates these arguments within the context of the progress of the Ghanaian Supreme Court in advancing a presumption of equitable ownership of property acquired during marriage and the spousal provisions in Ghana's Lands Act 2020 (Act 1036), which codify the presumption established by the Courts. The paper highlights social reproductive labour, public awareness, and the snail-paced legal system as structural, traditional, and historical issues that may impact Ghanaian women's right to property acquired during marriage. It argues that while legislation is vital in harmonising and providing a baseline of women's rights, law reforms alone are not enough to close the gender gap regarding women's matrimonial property rights. Stakeholders must recognise women's gendered position in these different contexts and how norms and values at different levels combine to situate their claim to resources. Such a holistic approach may successfully achieve an encompassing legal, regulatory, and social framework that safeguards women's marital property rights.

Keywords: joint ownership, women's marital property rights, African feminism, sustainable development goals, landed property rights

* PhD Candidate, University of Warwick; BA, LLB, LLM (BPP). I am grateful to Ann Stewart, Serena Natile, Maame AS Mensah-Bonsu, and Araba Nunoo for their comments on earlier drafts.

I. Introduction

Over the past ten years, inclusive capitalist discourses have promoted land privatisation and individual land ownership as a means of empowering women and fostering economic prosperity.[1] According to these narratives, equalitarian property rights for women impacts their well-being, decision-making ability in the home, and agency.[2] It also leads to an increase in children's welfare in a household.[3] The United Nations (UN) Sustainable Development Goals ('SDGs') follow this notion under Goal 5, which targets access to ownership and control over land and other forms of property to give women equal rights to economic resources.[4] Property acquired through marriage is one of the channels by which women are said to obtain access to property and one of the channels that require an adequate legal framework to protect the rights of married women.[5]

On this basis, international development policies and treaties have progressively sought to enhance women's rights to property acquired during a marriage (marital property). Marital property regimes are classified under four categories, according to researchers.[6] The first is 'full community of property', where all assets are regarded as the joint property of a spouse. The second option is 'partial community of property', which implies joint ownership of marital assets but still permits spouses to keep assets they obtained before marriage. The third is 'separation of property', in which every property, including those obtained through

[1] Abena D Oduro, Louis Boakye-Yiadom, and William Baah-Boateng, 'Asset Ownership and Egalitarian Decision-Making among Couples: Some Evidence from Ghana' (2012) 14 The Gender Asset Gap Project, Indian Institute of Management Bangalore Working Paper No 14, 9 <http://landwise-production.s3.amazonaws.com/2022/03/Oduro_Asset-ownership-and-egalitarian-decision-making-among-couples_2012.pdf> accessed 15 March 2023; Abena D Oduro, Carmen Diana Deere, and Zachary B Catanzarite, 'Women's Wealth and Intimate Partner Violence: Insights from Ecuador and Ghana' (2015) 21 Feminist Economics 1, 2–3; Isis Gaddis, Rahul Lahoti, and Hema Swaminathan, 'Women's Legal Rights and Gender Gaps in Property Ownership in Developing Countries' (2022) 48 Population and Development Review 331, 332.

[2] Oduro, Boakye-Yiadom, and Baah-Boateng (n 1); Oduro, Deere, and Catanzarite (n 1); Gaddis, Lahoti, and Swaminathan (n 1).

[3] Gaddis, Lahoti, and Swaminathan (n 1); Oduro, Deere, and Catanzarite (n 1); Cheryl Doss, 'The Effects of Intrahousehold Property Ownership on Expenditure Patterns in Ghana' (2006) 15 Journal of African Economies 149.

[4] United Nations, 'THE 17 GOALS' (*Sustainable Development Goals*) <https://sdgs.un.org/goals> accessed 18 April 2021.

[5] Gaddis, Lahoti, and Swaminathan (n 1); Oduro, Deere, and Catanzarite (n 1); Oduro, Boakye-Yiadom, and Baah-Boateng (n 1).

[6] Carmen Diana Deere and Cheryl R Doss, 'The Gender Asset Gap: What Do We Know and Why Does It Matter?' (2006) 12 Feminist Economics 1; Carmen Diana Deere and others, 'Property Rights and the Gender Distribution of Wealth in Ecuador, Ghana and India' (2013) 11 The Journal of Economic Inequality 249, 256–262; Gaddis, Lahoti, and Swaminathan (n 1).

marriage, is owned individually.[7] The last regime is known as a 'deferred community of property regime', in which property is considered privately owned until the marriage is dissolved; at this point, a full or partial community of property rule will apply.[8]

In Ghana, the issue of women's rights to co-own marital property has been a subject of debate and progressive judicial interpretation over the years. Marriage in Ghana is a contract between the man's and the woman's family and is surrounded by 'definite customs and laws' for the security of both parties.[9] This definition is in line with the Ghanaian court's emphasis that marriage is a union not just between the man and woman but also between their two families.[10] Although marriage joins the two families together, the man and woman remain members of their lineage and do not become entitled to any rights the other spouse can claim through their family.[11] Accordingly, Ghanaian women retain their legal identity after marriage as the woman's legal status is not subsumed into her husband's.[12]

Spouses maintaining their separate identities and individuality in marriage raises the question of whether couples must share whatever property they acquire during the marriage. The progressive steps towards protecting women's rights over matrimonial property have come from concerted efforts, through legislation and judicial reform, to thrust women's issues from the shadows into the forefront of national discourse. The current position of the Ghanaian Supreme Court is that there is a presumption of equitable ownership of property acquired during marriage.[13] In 2020, the Ghanaian parliament passed the Ghana Lands Act (Act 1036). The Act contained spousal provisions in sections 34 and 47, which legislate the presumption established by the Courts. Feminist activists and legal practitioners heralded the spousal provisions in the Act as a giant step in protecting women's rights to marital property.[14]

[7] World Bank Group, *Women, Business and the Law 2016: Getting to Equal* (World Bank 2015) <https://openknowledge.worldbank.org/handle/10986/22546> accessed 15 December 2022; Gaddis, Lahoti and Swaminathan (n 1).

[8] World Bank Group (n 7); Gaddis, Lahoti and Swaminathan (n 1).

[9] JWA Amoo, 'The Effect of Western Influence on Akan Marriage' (1946) 16 Africa 228, 228.

[10] *Yaotey v Quaye* [1961] GLR 573 (HC Ghana).

[11] Amoo (n 9) 228.

[12] *Acheampong v Acheampong* [1967] GLR 34 (HC Ghana); Gaddis, Lahoti and Swaminathan (n 1) 3.

[13] ; *Boafo v Boafo* [2005] SCGLR 705 (SC Ghana); *Gladys Mensah v Stephen Mensah* [2012] GHASC 8 (SC Ghana); *Quartson v Quartson* [2012] SCGLR 1077 (SC Ghana); *Patience Arthur v Moses Arthur (No 1)* Civil Appeal No J4/19/2013 (SC Ghana, 13 July 2013).

[14] JoyNews, 'Newsfile on Joy News' (*Youtube*, 10 April 2021) <https://www.youtube.com/watch?v=NiZtzJDqaq8&t=6950s> accessed 7 June 2021; ClearwayLaw Law School Series, 'The New Land Act Ghana' (*ClearwayLaw*, 28 February 2021) <https://clearwaylaw.com/the-new-land-act-ghana/> accessed 7 June 2021.

This article draws on feminist arguments that the law, as experienced by women, cannot be adequately explained by focusing exclusively on male-centric state law.[15] Feminist scholars argue that women belong to and are influenced by different social, economic, and political orders within a State.[16] Moore terms this as semi-autonomous social fields.[17] Ghana's legal regime on land rights is pluralistic and based on the 'co-existence of different regulatory systems, consisting of a hybrid of English common law principles, Ghanaian customary law principles, constitutional provisions, and statutory provisions.'[18] These laws are not mutually exclusive but overlap and interact in different contexts and represent women's social, political, and economic constructions.[19] Therefore, understanding how laws affect women and bringing about social change that improves women's access to landed property cannot be done by evaluating the rights guaranteed by State law in isolation.[20] The paper situates these arguments within the context of the progress of the Ghanaian Supreme Court in advancing a presumption of equitable property ownership and the provisions in the Ghana Lands Act 2020 codifying the judicial precedent. This research draws on data from primary legal sources— legislation, case law, and non-legal sources—radio and television interviews.[21] Cases selected are precedent-setting cases resolved after the coming into force of the 1992 Constitution.[22]

This paper adds to the body of literature on women's property rights in Ghana. Extant literature has examined inheritance laws and their implications on women's property rights.[23] Women's property rights after divorce have also been

[15] Ambreena S Manji, 'Imagining Women's "Legal World": Towards a Feminist Theory of Legal Pluralism in Africa' (1999) 8 Social & Legal Studies 435, 450; Anne Hellum and others, *Human Rights, Plural Legalities, and Gendered Realities: Paths Are Made by Walking* (SEARCWL with Weaver Press 2007).

[16] Manji (n 15) 450; Hellum and others (n 15); Anne MO Griffiths, 'Making Gender Visible in Law: Kwena Women's Access to Power and Resources' in Anne Hellum and others (eds), *Human Rights, Plural Legalities, and Gendered Realities: Paths are Made by Walking* (SEARCWL with Weaver Press 2007) 139; Ambreena Manji and Ann Stewart, 'I Built This House on My Back' in Sam Adelman and Abdul Paliwala (eds), *Beyond Law and Development: Resistance, Empowerment and Social Injustice* (1st edn, Routledge 2022).

[17] Sally Falk Moore, 'Law and Social Change: The Semi-Autonomous Social Field as an Appropriate Subject of Study' (1973) 7 Law & Society Review 719.

[18] Lennox Agbosu and others, *Customary and Statutory Land Tenure, and Land Policy in Ghana: Paper Prepared for the Land Tenure and Land Policy Research Project* (Institute of Statistical, Social & Economic Research, University of Ghana 2007) 32–33.

[19] Ann Stewart, *Gender, Law and Justice in a Global Market* (Cambridge University Press 2011) 59–60.

[20] Manji (n 15); Griffiths (n 16); Griffiths, *Transformations on the Ground: Space and the Power of Land in Botswana* (Indiana University Press 2019) 30.

[21] Alan Bryman, *Social Research Methods* (Oxford University Press 2016) 183.

[22] See *Peter Adjei v Margaret Adjei* [2021] GHASC 5 (SC Ghana) 9.

[23] Henrietta JAN Mensa-Bonsu, 'The Intestate Succession Law of Ghana: Practical Problems in Application' (1994) 8 Jahrbuch für afrikanisches Recht 105; C Dowuona-Hammond, 'Ensuring Equity in the

the subject of other research.[24] A few studies have combined women's legal and traditional property rights in urban and rural locations.[25] By addressing the historical and structural reasons that affect women's property rights and result in latent gender biases in legislation, this study adds to the body of work in this area.

First, the paper discusses the discourses on co-ownership of marital property as a tool to promote women's economic rights and the international law regime protecting these rights. Second, it discusses Ghana's regime on landed property, and looks briefly at the Ghanaian Supreme Court's progression to equitable right to marital property and the sections in the Ghana Lands Act, which codify some aspects of the Court's precedent. Finally, the paper highlights social reproductive labour, public awareness, and the snail-paced legal system of Ghana as the structural, traditional, and historical issues that may impact Ghanaian women's right to property acquired during marriage. It argues that while legislation is vital in harmonising and providing a baseline of women's rights, law reforms alone are not enough to close the gender gap regarding women's matrimonial property rights. Stakeholders must recognise women's gendered position in these different contexts and how norms and values at different levels combine to situate their claim to resources. Such a holistic approach may be successful in a holistic approach to achieving an encompassing legal, regulatory, and social framework that safeguards women's marital property rights.[26]

Distribution of Matrimonial Property upon Divorce: Preparing the Path for Legislation' (2005) 2 University of Botswana Law Journal 101; Akua Kuenyehia, 'Women, Marriage, and Intestate Succession in the Context of Legal Pluralism in Africa' (2006) 40 UC Davis Law Review 385; Takyiwaa Manuh, 'Wives, Children, and Intestate Succession in Ghana' in Gwendolyn Mikell (ed), *African Feminism* (University of Pennsylvania Press 2010) 77; Ama Hammond, 'Reforming the Law of Intestate Succession in a Legally Plural Ghana' (2019) 51 The Journal of Legal Pluralism and Unofficial Law 114.

[24] Dowuona-Hammond (n 23); Ama Fowa Hammond, 'What Man Has Put Together-Recognising Property Rights of Spouses in De Facto Unions' (2008) 24 University of Ghana Law Journal 231; Henrietta JAN Mensa-Bonsu, 'Ensuring Equitable Access to Marital Property When the Holy Estate Becomes an Unholy Ex-State: Will the Legislature Walk the Road Paved by the Courts' (2011) 25 University of Ghana Law Journal 99; Maame Yaa Mensa-Bonsu and Maame AS Mensa-Bonsu, 'To Win Both the Battle and the War: Judicial Determination of Property Rights of Spouses in Ghana' in Josephine Jarpa Dawuni *Gender* (ed), *Judging and the Courts in Africa* (1st edn, Routledge 2021).

[25] Ellen Bortei-Doku Aryeetey, 'Behind the Norms: Women's Access to Land in Ghana' in C Toulmin, Deville P Lavigne, and S Traoré (eds), *Managing Land Tenure and Resource Access in West Africa* (James Currey Ltd 1997); Christine Dowuona-Hammond, 'State Land Management Regime: Impact on Land Rights of Women and the Poor in Ghana' (GTZ Legal Pluralism and Gender Project 2003); S Minkah-Premo and C Dowuona-Hammond, 'Review of Land and Gender Studies and Identification of Resources in Ghana' (Ghana Land Administration Project 2005); Deere and others (n 6); Akua O Britwum and others, *Gender and Land Tenure in Ghana: A Synthesis of the Literature* (ISSER, Institute of Statistical, Social & Economic Research, University of Ghana 2014); Sheila Minkah-Premo, 'Report Of Desktop Study On Gender Equality And Social Inclusion Issues Relating To The Land Sector In Ghana' (NETRIGHT 2018) <https://www.star-ghana.org/learning-2/publications-and-resources/reports/137-netright-study-on-gender-equality-and-social-inclusion-issues-relating-to-the-land-sector-in-ghana/file> accessed 18 May 2021.

[26] Griffiths (n 16) 139.

II. Joint Ownership of Marital Property: A Pathway To Property Ownership

Extant literature points to the importance of legal property rights for women.[27] According to development policies, land title formalisation and individual land ownership in developing countries promote economic growth and poverty alleviation.[28] Development policies identify property rights as critical for achieving gender equality, women's empowerment and bridging the financial gap for women globally.[29] The World Bank reports that 'resources, agency, and achievements' are three central components of empowerment.[30] Agency in this context refers to people's ability to act upon plans and lead the lives they desire.[31] Researchers who link women's empowerment with agency argue that the two have an interdependent relationship.[32] Empowering women increases agency, which increases the likelihood of successful development outcomes, and vice versa.[33] The World Bank's conceptualisation of agency is similar to Kabeer's conceptualisation. Kabeer points to agency as the second pathway under which empowerment emerges.[34] Kabeer links agency with the condition of choice, where the person can appraise and choose options or pre-conditions that suit them. The rationale for the link between agency and empowerment is that when women are given external resources, such as land and credit and internal resources, like self-confidence and knowledge, they can enjoy integrity, bodily autonomy, and personal freedoms.[35] Research finds that women's participation in household decision-making increases

[27] Oduro, Boakye-Yiadom, and Baah-Boateng (n 1); Oduro, Deere, and Catanzarite (n 1); Gaddis, Lahoti, and Swaminathan (n 1).

[27] Dowuona-Hammond (n 23); Doss (n 3); Mensa-Bonsu, 'What Man Has Put Together-Recognising Property Rights of Spouses in De Facto Unions' (n 24); Oduro, Deere, and Catanzarite (n 1); Hammond (n 23); Mensa-Bonsu and Mensa-Bonsu (n 24); Gaddis, Lahoti and Swaminathan (n 1).

[28] See for example United Nations (n 4).

[29] ibid.

[30] World Bank Group, 'Profiting from Parity: Unlocking the Potential of Women's Business in Africa' (World Bank 2019) 38 <https://openknowledge.worldbank.org/handle/10986/31421> accessed 15 December 2022.

[31] Gita Sen and Avanti Mukherjee, 'No Empowerment without Rights, No Rights without Politics: Gender-Equality, MDGs and the Post-2015 Development Agenda' (2014) 15 Journal of Human Development and Capabilities 188.

[32] Naila Kabeer, 'Resources, Agency, Achievements: Reflections on the Measurement of Women's Empowerment' (1999) 30 Development and Change 435, 457–461; Sen and Mukherjee (n 31); Bougema Theodore Ntenkeh, Dobdinga Cletus Fonchamnyo and Denis Nfor Yuni, 'Women's Empowerment and Food Security in Cameroon' (2022) 56 The Journal of Developing Areas 141.

[33] Kabeer, 'Resources, Agency, Achievements' (n 32); Sen and Mukherjee (n 31); Ntenkeh, Fonchamnyo, and Yuni (n 32).

[34] Naila Kabeer, 'Gender Equality and Women's Empowerment: A Critical Analysis of the Third Millennium Development Goal 1' (2005) 13(1) Gender & Development 13.

[35] Sen and Mukherjee (n 31) 190.

when marital property is distributed equally.[36] It has been argued that women's equal access to property within marriage gives them more substantial bargaining power as they have options outside of the marriage.[37] Similarly, other research finds that women have more security and are less likely to suffer violence when they are not structurally dependent on men for access to resources.[38] Therefore, joint ownership of marital property is linked to empowering women by improving their agency.

On the international front, the Protocol to the African Charter on Human and Peoples' Rights on the Rights of Women in Africa ('Maputo Protocol')[39] and the Convention on the Elimination of All Forms of Discrimination Against Women ('CEDAW')[40] provide overlapping and yet divergent protection for the right to joint ownership of marital property. CEDAW and the Maputo Protocol place the duty on State Parties to use 'appropriate legislative, institutional and other measures to fight against all forms of discrimination against women'.[41] Both treaties provide equal rights to property acquired during marriage. Article 16 of CEDAW provides that state parties are to ensure the elimination of discrimination against women 'in all matters relating to marriage and family relations' and to ensure that women have the same rights in 'respect of the ownership, acquisition, management, administration, enjoyment and disposition of property, whether free of charge or for a valuable consideration. On the regional level, the Maputo Protocol brings specificity to the African reality and places CEDAW within the context of the issues that concern African women.[42] The Protocol emphasises under article 2 that 'positive action' is required to ensure the realisation of women's rights within the private domain and recognises the precarious situation that married women in Africa face regarding equal rights to matrimonial property upon divorce or death intestate of their spouses. Significantly, the Protocol recognises a right to culture while challenging and weeding out negative norms and practices justified through culture. 'Positive African cultural values' are defined within the Protocol as values 'based on the principles of equality, peace, freedom, dignity, justice,

[36] Oduro, Boakye-Yiadom, and Baah-Boateng (n 1) 18; Gaddis, Lahoti, and Swaminathan (n 1).
[37] Oduro, Boakye-Yiadom, and Baah-Boateng (n 1); Gaddis, Lahoti, and Swaminathan (n 1).
[38] Minkah-Premo (n 25) 41.
[39] Protocol to the African Charter on Human and Peoples' Rights on the Rights of Women in Africa 2003 (adopted 1 July 2003, entered info force 25 November 2005) ('Maputo Protocol') <www.ohchr.org/sites/default/files/Documents/Issues/Women/WG/ProtocolontheRightsofWomen.pdf> accessed 19 March 2023.
[40] Convention on the Elimination of All Forms of Discrimination against Women (adopted 18 December 1979, entered into force 3 September 1981) 1249 UNTS 13 ('CEDAW').
[41] CEDAW, art 2(1).
[42] Frans Viljoen, 'An Introduction to the Protocol to the African Charter on Human and Peoples' Rights on the Rights of Women in Africa' (2009) 16 Washington and Lee Journal of Civil Rights and Social Justice 11, 22.

solidarity, and democracy'.[43] It might be argued that the Protocol's structure suggests a restriction on cultural plurality because it only acknowledges cultural values that are regarded to be founded on the standards it established. Nonetheless, the acknowledgement of culture within the framing of the Protocol symbolises a recognition that African women do not have to 'strip themselves' of their culture and forego their identity before they can access their rights.[44] As Musembi argues, policymakers should not view 'culture' and 'rights' as 'polar opposites', where culture is perceived 'largely as a negative force that impedes the realisation of rights.'[45] Women have, in some instances, used culture as a medium to assert their rights.[46] Therefore, any legislation on women's rights must acknowledge the influence of culture and social norms on women's life.[47]

Within the current inclusive capitalist policy discourse, the right to equal co-ownership of marital property is addressed under the SDGs adopted by the UN member organisations.[48] Goal 5 includes '…empower all women and girls' as its gender equality objective.[49] To emphasise the importance of 'empowerment' in the SDGs, some researchers assert that even though women's empowerment is stated in Goal 5, women's empowerment is a crucial element in accomplishing all the other SDGs.[50] Target 5a focuses on access to ownership and control over land and other forms of property to give women equal rights to economic resources.[51] This target is measured by 'the proportion of countries where the legal framework (including customary law) guarantees women's equal rights to land ownership and/or control'.[52] The SDGs targets do not specifically mention joint ownership of marital property. Nonetheless, subsequent international financial institution reports include equal ownership of marriage as a life event and equal rights to property

[43] Viljoen (n 42) 18.
[44] Celestine Nyamu Musembi, 'Pulling Apart? Treatment Of Pluralism In CEDAW And Maputo Protocol' in Anne Hellum and Henriette Sinding Aasen (eds), *Women's Human Rights : CEDAW in International, Regional, and National Law* (Cambridge University Press 2013) 204–205. See also Celestine I Nyamu, 'How Should Human Rights and Development Respond to Cultural Legitimization of Gender Hierarchy in Developing Countries' (2000) 41 Harvard International Law Journal 381, 391–393.
[45] Musembi (n 44) 204–205. See also Nyamu (n 44) 391–393.
[46] Abdullahi Ahmed An-Na'im, 'Introduction' in Abdullahi Ahmed An-Na'im and Jeffrey Hammond (eds), *Cultural Transformation and Human Rights in Africa* (Zed Books Ltd 2002); Celestine Nyamu-Musembi, 'Are Local Norms and Practices Fences or Pathways: The Example of Women's Property Rights' in Abdullahi Ahmed An-Na'im and Jeffrey Hammond (eds), *Cultural Transformation and Human Rights in Africa* (Zed Books 2002).
[47] Musembi (n 44) 203–204.
[48] United Nations (n 4).
[49] ibid.
[50] Jummai Othniel Yila and Almamy Sylla, 'Women Empowerment in Addressing Food Security and Nutrition' in Walter Leal Filho and others (eds), *Zero Hunger* (Springer 2020); Ntenkeh, Fonchamnyo and Yuni (n 32).
[51] United Nations (n 4).
[52] ibid.

acquired during marriage as one of the three pathways for women to acquire land.[53]

Ghana ratified the Maputo Protocol in 2007 without reservation.[54] However, the country does not have explicit laws protecting the joint ownership of marital property. The following section briefly discusses Ghana's current regime on co-ownership and the structural and historical challenges that make solely focusing on legislation inadequate to address women's matrimonial property rights.

III. Approaching The Issue Of Women's Marital Property Rights Holistically

In the preceding section, this paper highlighted the arguments that equal ownership of property acquired during a marriage is significant for enhancing women's empowerment and economic growth. Joint ownership of marital property is argued to increase women's household decision-making, strengthen their bargaining power, and release economic resources that all combine to enhance women's economic growth. This section briefly discusses Ghana's current regime on marital property.

The responsibility of adapting the legislation pertaining to women's rights to marital property to fit societal changes has fallen on the shoulders of the Ghanaian Supreme Court.[55] In doing this, it has relied predominantly on the entrenchment of the right to equal opportunities and people's social, economic, and cultural rights regardless of gender under Ghana's 1992 Constitution[56] and the automatic enforceability of human rights guaranteed by international treaties in the Ghanaian Courts.[57] The Court has moved steadily from its initial approach, where women had no claim to matrimonial property, to equal ownership of the same.[58] This approach reinforced the traditional hierarchical family structure, with the man as the head and the woman and children as subordinates.[59] The Court currently advances the 'equality is equity principle'. The Court has asserted that where the spouses have no contrary agreement, any property acquired during the subsistence of the marriage is deemed joint property to be shared equally on

[53] Oduro, Boakye-Yiadom, and Baah-Boateng (n 1); Gaddis, Lahoti, and Swaminathan (n 1).
[54] Executive Council of the African Union, 'Report on the Status of OAU/AU Treaties' (EX.CL/728(XXI), African Union 11 July 212) <www.peaceau.org/uploads/ex-cl-728-xxi-e.pdf> accessed 7 April 2020.
[55] Mensa-Bonsu and Mensa-Bonsu (n 24); Dowuona-Hammond (n 23); Mensa-Bonsu, 'What Man Has Put Together-Recognising Property Rights of Spouses in De Facto Unions' (n 24).
[56] Constitution of the Republic of Ghana 1992, art 12.
[57] Constitution of the Republic of Ghana 1992, art 33(5).
[58] *Quartey v Martey* [1959] GLR 377 (HC Ghana).
[59] Stefano Boni, 'The Encompassment of the Autonomous Wife: Hierarchy in Akan Marriage (Ghana)' (2002) 97 Anthropos 55.

divorce.[60] The basis of this decision was that ordinary incidents of commerce have no application in marital relations between husband and wife who jointly acquire property during the marriage.[61] Despite this, the Court has been mindful of respecting the fundamental right given to each person under the Constitution to own property solely.[62] Nevertheless, the equitable maxim of 'equality is equity' does not mean that the Court favoured equal sharing of joint property in all circumstances.[63] The Court determines what is 'equitable' and the proportions each spouse is entitled to, purely on a case-by-case basis.[64]

Apart from the provisions within the 1992 constitution, the Lands Act[65] passed in 2020 is heralded as one of the legislative interventions protecting women's right to matrimonial property.[66] The Act legislates the principle established by the Ghanaian Supreme Court that unless the spouses express a contrary intention, there is a presumption favouring equitable ownership of all properties acquired during marriage. In that respect, the law will deem spouses to be parties in conveying an interest in land. Thus, the presumption will apply unless the spouses expressly disclose on the face of the conveyance that the property being registered belongs solely to one spouse.[67] Although the Registrar of Lands will only register the interest in land in the joint name of both spouses,[68] registration in only one spouse's name no longer poses an issue. The spouse with the legal title holds the property in trust for themselves and the other spouse.[69] The combined effect of these provisions is that a spouse cannot unilaterally transfer the interest of landed property through sale, exchange, lease, or mortgage, without the other spouse's express agreement.

From the above discussion, it can be inferred that Ghana follows the 'partial community' of property regime. Although the Ghanaian Constitution does not define what constitutes a 'jointly acquired property', the Ghanaian Supreme Court has formulated a working definition. According to the Court, where the spouses have no contrary agreement, any property acquired during the subsistence of the marriage is deemed joint property to be shared equally on divorce.[70] Thus property acquired before marriage does not count as marital property. By presuming that any landed property acquired during marriage belongs to both spouses, this law protects an individual's right to own property while ensuring the protection

[60] *Mensah v Mensah* (n 13).
[61] See also *Patience Arthur v Moses Arthur (No 1)* (n 13) 559.
[62] Constitution of the Republic of Ghana 1992, art 18.
[63] *Boafo v Boafo* (n 13).
[64] ibid (Dr Date-Bah JSC).
[65] Ghana Land Act 2020 (Act 1036).
[66] JoyNews (n 14); ClearwayLaw Law School Series (n 14).
[67] ClearwayLaw Law School Series (n 14); Ghana Land Act 2020, s 47.
[68] Ghana Land Act 2020, s 97(4).
[69] Ghana Land Act 2020, s 38(4).
[70] See also *Patience Arthur v Moses Arthur (No 1)* (n 13) 559.

and valuation of the labour that spouses put into the development and sustenance of the family.

In sum, Ghanaian women are protected in different planes by international and regional treaties and national laws that protect the property rights of all citizens and the particular laws that specifically protect women's property rights. While further legislation is needed, particularly a revised Matrimonial Causes Act to deal with the specificities of the distribution of property upon divorce, that discussion is beyond the scope of this article. The final sections of this paper focus on the impact that social reproductive labour, the snail-paced Ghanaian court system, and sociocultural norms, such as polygamy, have on women's marital property rights.

A. Social reproductive labour as non-monetary contributions

Under colonialism, the law significantly contributed to the construction of gender relations with differentiation between male and female labour and a devaluation of women's labour.[71] The 'patriarchal coalition'[72] formed between the colonial government and the traditional chiefs and elders aimed at socially controlling women[73] further ingrained the existing inequality between African men and women.[74] Women were excluded from men's economic activities as men were engaged in meeting their colonial tax obligations. The men's work in mines, commercial farms, and construction was characterised as in the 'public sphere' and was connoted with economic value.[75] In contrast, women controlling agricultural produce and performing reproductive and social care were connoted in the 'private' sphere and had no monetary value.[76] These colonial gender constructions

[71] Stewart, *Gender, Law and Justice in a Global Market* (n 19) 105.
[72] Jane L Parpart and Kathleen A Staudt, 'Women and the State in Africa' in Jane L and Staudt Parpart (eds), *Women and the State in Africa* (Lynne Rienner Publishers 1989).
[73] Martin Chanock, 'Making Customary Law: Men, Women, and Courts in Colonial Northern Rhodesia' in Margaret Jean Hay and Marcia Wright (eds), *African Women and the Law: Historical Perspectives* (Boston University Papers on Africa 1982) 67; Manji (n 15) 445.
[74] Margot L Lovett, 'Gender Relations Class Formation and the Colonial State in Africa' in Jane L and Staudt Parpart (eds), *Women and the State in Africa* (Lynne Rienner Publishers 1989) 25; Parpart and Staudt (n 72); Chanock (n 73).
[75] Lovett (n 74); Marjorie J Mbilinyi, '"This Is an Unforgettable Business": Colonial State Intervention in Urban Tanzania' in Jane L and Staudt Parpart (ed), *Women and the State in Africa* (Lynne Rienner Publishers 1989); Catherine Marshall and Gary L Anderson, 'Rethinking the Public and Private Spheres: Feminist and Cultural Studies Perspectives on the Politics of Education' (1994) 9 Journal of Education Policy 169.
[76] Stewart, *Gender, Law and Justice in a Global Market* (n 19) 101.

continue to resonate in post-colonial Africa.[77] Presently, African women's property rights and consequent economic status continues to be affected by 'discriminatory laws, cultural and/or religious norms, and traditions that perpetuate their exclusion from access, and control over resources'.[78] Gayoye captures this position clearly when she asserts that African women's exclusion from access to property has a colonial legacy that has had a lasting and stubborn impact.[79]

Ghana's property rights of women in marriage do not follow the British colonial doctrine of 'couverture'. Under the doctrine of coverture, the legal status of women was subsumed into their husbands' upon marriage.[80] The Ghanaian Courts have maintained the customary law position that there is no 'legal fiction' in a marriage that a husband and wife are one in law.[81] Both parties retain their separate identities and individuality.[82] This customary separation of spouses as separate entities is arguably one of the starting points of the issues women face in asserting their rights to matrimonial property. Where there is no legal fiction of oneness, the Courts have had to determine whether property acquired during a marriage is jointly owned.

Gender gaps in economic opportunities and earnings may impact women's ability to acquire market-based property.[83] Ghanaian women's increased participation in the formal and informal sectors[84] of the economy does not mean they no longer perform socially productive roles of domestic chores, childbearing, and child-raising, which are traditionally ascribed to women within the family. Studies show that while women increasingly join the labour force in formal or informal employment, women still bear the traditional responsibility of social reproductive work.[85] Studies also show that married women are more likely to have a higher

[77] Sally Engle Merry, 'From Law and Colonialism to Law and Globalization' (2003) 28 Law & Social Inquiry 569, 582.

[78] Meskerem Geset Techane, 'Economic Equality and Female Marginalisation in the SDGs Era: Reflections on Economic Rights of Women in Africa' (2017) 1 Peace Human Rights Governance 333.

[79] Martha Gayoye, 'Why Women Judges Really Matter: The Impact of Women Judges on Property Law Outcomes in Kenya' (2021) 31 Social & Legal Studies 72.

[80] Gaddis, Lahoti and Swaminathan (n 1); Deere and Doss (n 6).

[81] *Acheampong v Acheampong* (n 12).

[82] Amoo (n 9).

[83] World Bank Group, *World Development Report 2012: Gender Equality and Development* (World Bank 2011) <https://openknowledge.worldbank.org/handle/10986/4391> accessed 15 December 2022; Gaddis, Lahoti and Swaminathan (n 1).

[84] Loretta Agyemang, 'Women's Work: Labor Market Outcomes and Female Entrepreneurship in Ghana' (2017) 3 Butler Journal of Undergraduate Research 2, 8.

[85] Janet Momsen, *Gender and Development* (2nd edn, Routledge 2009) 3; Clara Osei-Boateng and Edward Ampratwum, 'The Informal Sector in Ghana' (Friedrich-Ebert-Stiftung 2011) <https://library.fes.de/pdf-files/bueros/ghana/10496.pdf> accessed 21 March 2023; Gaëlle Ferrant, Luca Maria Pesando and Keiko

purchasing power if both spouses work.[86] Nevertheless, research finds that the urban married woman's purchasing ability is negatively affected by her work.[87] Studies have offered various explanations for this discrepancy in the impact of women's employment on their purchasing power. According to some studies, women's unpaid work may prevent them from entering the workforce and lowers their earning potential, particularly in their reproductive stages.[88] Other research finds that women's unpaid labour constrains the choice of paid work available to them both in the formal and informal sectors of the economy.[89] Although studies find that social reproductive labour plays a vital role in the State economy and household sustenance,[90] this labour continues to be unpaid and economically devalued.[91] As a result, although the Ghanaian family has largely moved away from the traditional idea of the man being the family's only earner, women's labour is increased, undervalued, and generates less revenue. The result is that married women may suffer the most negative impact if legislation and case law do not adequately address non-monetary contributions.[92]

The Ghanaian Supreme Court, in recent years, recognises that a spouse undertaking the household's various tasks to relieve the partner and enable them to acquire property is a contribution to the marriage.[93] Nonetheless, such

Nowacka, 'Unpaid Care Work: The Missing Link in the Analysis of Gender Gaps in Labour Outcomes' (OECD December 2014) <https://www.oecd.org/dev/development-gender/Unpaid_care_work.pdf> accessed 15 December 2022; United Nations, 'UN Women Annual Report 2015–2016' (UN Women 2016) <https://annualreport.unwomen.org/en/2016> accessed 12 September 2022; Mariama Awumbila, Joseph Kofi Teye, and Joseph Awetori Yaro, 'Of Silent Maids, Skilled Gardeners and Careful Madams: Gendered Dynamics and Strategies of Migrant Domestic Workers in Accra, Ghana' (2017) 82 GeoJournal 957; Mariama Awumbila and others, 'Please, Thank You and Sorry–Brokering Migration and Constructing Identities for Domestic Work in Ghana' (2019) 45 Journal of Ethnic and Migration Studies 2655; Manji and Stewart, 'I Built This House on My Back' (n 16) 104.

[86] Gaddis, Lahoti, and Swaminathan (n 1) 7.

[87] Gaddis, Lahoti, and Swaminathan (n 1) 7.

[88] Naila Kabeer, 'Gender Equality, Inclusive Growth, and Labour Markets' in Kate Garantham, Gillian Dowie and Arjan de Haan (eds), *Women's Economic Empowerment: Insights from Africa and South Asia* (Routledge 2021) 21.

[89] Shahra Razavi, 'The Political and Social Economy of Care in a Development Context: Contextual Issues, Research Questions and Policy Options' (Gender and Development Programme Paper No 3, UNRISD 2007) <https://cdn.unrisd.org/assets/library/papers/pdf-files/razavi-paper.pdf> accessed 21 March 2023; Ferrant, Pesando and Nowacka (n 85); Merike Blofield and Merita Jokela, 'Paid Domestic Work and the Struggles of Care Workers in Latin America' (2018) 66 Current Sociology 531.

[90] Diane Elson, 'Labor Markets as Gendered Institutions: Equality, Efficiency and Empowerment Issues' (1999) 27 World Development 611.

[91] Catherine Hoskyns and Shirin M Rai, 'Recasting the Global Political Economy: Counting Women's Unpaid Work' (2007) 12 New Political Economy 297, 306; Stewart, *Gender, Law and Justice in a Global Market* (n 19) 16.

[92] Gaddis, Lahoti, and Swaminathan (n 1).

[93] *Boafo v Boafo* (n 13); *Gladys Mensah v Stephen Mensah* (n 13) 12; *Peter Adjei v Margaret Adjei* (n 22) 23. See also *Beauty Katey v William Kwadwo Katey* Suit No H1/176/2015 (CA Ghana, 14 July 2016) 6

contributions are seen as non-pecuniary/non-monetary because no financial value is attached to social reproductive labour. The court accepts the woman's role as a substantive contribution to property acquired during marriage because it gives the spouse the 'free hand to engage in economic activities.'[94] Thus, to the courts, social reproductive labour is not an economic contribution. It is viewed as a contribution because of the help it offers to the partner to pursue what the court sees as actual economic activity. According to the Court's reasoning,

> a partner who performs various household chores for the other… such that the other partner has a free hand to engage in economic activities must not be discriminated against in the distribution of properties acquired during the marriage when the marriage is dissolved. This is so because it can safely be argued that the properties' acquisition was facilitated by the massive' assistance that the other spouse derived from the other.[95]

Again, joint ownership does not mean equal ownership of marital property under the Ghanaian regime. According to the Supreme Court, the equitable maxim of 'equality is equity' the courts apply does not mean equal ownership. To the Court, 'the question of what is "equitable", in essence, what is just, reasonable and accords with common sense and fair play, is a pure question of fact, dependent purely on the particular circumstances of each case'.[96] Consequently, to avoid manifest injustice to one party, the issue of proportions would be dealt with in accordance with the equities of each case and not by a blanket application.[97]

In the recent case of *Peter Ajei v Magaret Adjei*, the Supreme Court expressed that it is a 'judicially created presumption' that properties acquired during a marriage are the joint property of the spouses, which does not confer substantive rights.[98] The Court's treatment of the rights to joint ownership of marital property as a non-substantive right means that judges can weigh the social reproductive work and decide if it is beyond what is customarily expected of a spouse within a marriage. With this approach, judges have an unfettered free hand to determine how much value can be attached to social reproductive work. In the same case, the Court decided that a wife could not be held to be a joint owner because 'seriously speaking', she did not give any indication of the husband's work, how he made his earnings and acquired the properties.[99] Accordingly, the court decided that she was not entitled to the marital residence or the assets acquired because

[94] *Gladys Mensah v Stephen Mensah* (n 13) 7.
[95] ibid 8.
[96] ibid 13; *Mensah v Mensah* [1993] 1 GLR 111 (CA Ghana) 714.
[97] *Gladys Mensah v Stephen Mensah* (n 13) 13; *Mensah v Mensah* (n 93) 714.
[98] *Peter Adjei v Margaret Adjei* (n 22) 12.
[99] ibid 13.

she was unaware of her spouse's business dealings.[100] The wife's contribution to the family through her social reproductive labour was discounted as unimportant and did not substantially contribute to purchasing the properties in question.[101] Consequently, the Court holds that even while a wife is not professionally active in the same occupation as her husband, she is nonetheless obliged to be fully informed of his business affairs in addition to engaging in social reproductive work or other economic activities. This added burden that the Court has placed on women's right to marital property undertones an implicit bias and treatment of social reproductive labour as non-economic contributions to the household. With this reasoning, the Court reinforces the deep-seated belief that social reproductive work alone, regardless of its extent, is insufficient.

The Supreme Court's method of treating questions of marital property on a case-by-case basis is not unreasonable since it gives the courts room to take the intricacies of each case into account. However, it is crucial that the courts assign economic value to women's social reproductive labour rather than seeing it as non-pecuniary contributions. Deere and Doss argue that giving non-pecuniary value to women's social reproductive labour reinforces the gender economic gaps[102]As such, it is essential that social reproductive work is assigned an economic value in determining marital property matters. This approach will require legislative backing and a well-researched guidelines framework. It may present some early challenges because States do not currently account for domestic work when calculating their Gross Domestic Product (GDP).[103] Nonetheless, nations and international bodies have, in recent years, recognised the economic value of domestic work.[104] Currently, Ghana's Labour laws provide safeguards for those who conduct domestic work as a kind of employment, recognising them as workers.[105] Consequently, with further research, Ghana can develop guidelines to calculate the economic value of the social reproductive labour of spouses in a just and equitable manner.

[100] ibid.
[101] ibid.
[102] Deere and Doss (n 6) 9.
[103] Diane Elson, 'The Economic, the Political and the Domestic: Businesses, States and Households in the Organisation of Production' (1998) 3 New Political Economy 189, 202; Hoskyns and Rai (n 91) 299–300.
[104] Elson (n 103); Blofield and Jokela (n 89); Awumbila and others (n 85); UNICEF Office of Research Innocenti, 'Gender, Paid Domestic Work and Social Protection. Exploring Opportunities and Challenges to Extending Social Protection Coverage among Paid Domestic Workers in Nigeria' (*UNICEF-IRC*, 20 January 2020) <www.unicef-irc.org/article/1961-gender-paid-domestic-work-and-social-protection.html > accessed 20 March 2022.
[105] Labour (Domestic Workers') Regulations (LI 2408) 2020.

B. THE SNAIL-PACED LITIGATION SYSTEM

For the country's civil litigation system to be improved in terms of women's property rights, it must be shorter and less expensive. Unfortunately, myriad obstacles swarm the Ghanaian litigation system and hinder women's full participation in the justice system. The issue of the length of litigation is not peculiar to only women. In describing the Ghanaian litigation system, Atuguba asserts that:

> Ghana's justice system is not just slow. It is also expensive and sometimes even harsh. Initiation of an action through the wrong processes or a lack of funds to sustain even the most legitimate cases could mean that a person is denied their day in court. This could explain why some frustrated people are seeking justice outside the law.[106]

The high financial burden and cost of time attached to seeking justice in the unfair distribution of matrimonial property further widens the economic gap between the sexes and diminishes gains in protecting women's marital property rights.[107] This is particularly so in cases where the women are economically unequal to their spouses and inhabit vulnerable positions within marriages.[108] While it is true that Ghana's 1992 constitution, the Lands Act, the 'equality is equity' principle of the Courts, and the international instruments protect women's matrimonial property rights, progress can only be felt where justice is easy and speedy to access. The length of time and the costs associated with asserting their rights may deter women and hinder them from fully participating in the justice system.

For instance, the women who have brought cases to the Ghanaian courts for review were in a less favourable economic position than their spouses.[109] In the cases of *Arthur*[110] and *Quartson*[111], the claimant wives were homemakers, while the husbands were a sailor and an international footballer, respectively. In Arthur, the wife alleged she had given up participating in economic activity to be a homemaker and her husband's driver because he could not drive.[112] The husband in *Katey v Katey*[113] was a geodetic engineer, while his claimant wife was a beautician.

[106] Raymond Atuguba, 'Ghana's Justice System Needs a Major Overhaul: Here's What Should Be Done' (*The Conversation*, 4 January 2018) <https://theconversation.com/ghanas-justice-system-needs-a-major-overhaul-heres-what-should-be-done-88724> accessed 8 June 2021.
[107] Deere and others (n 6); Mensa-Bonsu and Mensa-Bonsu (n 24) 46–47.
[108] Deere and others (n 6); Mensa-Bonsu and Mensa-Bonsu (n 24) 46–47.
[109] Mensa-Bonsu and Mensa-Bonsu (n 24) 46.
[110] *Patience Arthur v Moses Arthur (No 1)* (n 13).
[111] *Quartson v Quartson* (n 13) 14.
[112] *Patience Arthur v Moses Arthur (No 1)* (n 13) 2.
[113] *Beauty Katey v William Kwadwo Katey* (n 93) 1.

Regarding the unbalanced economic positions of the parties in *Katey v Katey*, Acquaye JA stated, 'all things being equal, the ability of a geodetic engineer to earn income will be greater than that of a beautician selling second-hand clothes.[114] While their counterparts in the Supreme Court in England and Wales have taken a maximum of two years to resolve cases on matrimonial causes,[115] the Ghanaian Courts have taken between six and sixteen years to resolve similar cases.[116] While there may be other reasons for the delays in the judicial processes in Ghana, the inherent inefficiencies in Ghana's judicial system contribute to the length of litigation.[117] The wheels of justice grinding slowly in cases where the woman may already be economically and socially vulnerable does not augur well with protecting women's matrimonial property rights. Legal protections favouring women's equality are useless if women cannot easily access the courts or choose not to pursue litigation because they cannot afford the cost, do not trust the court system, or are frustrated by the amount of time it takes for cases in the court system to be determined. There must be an administrative, infrastructure and technical reform of the court system. Practical strategies, such as e-justice systems, and specialised women's courts, may improve the pace of the litigation process and reassure women of gender justice.

C. Sociocultural norms that affect the efficacy of laws

The task of regulating the property rights of spouses and entitlements in the event of dissolution has proven to be very challenging.[118] This, perhaps, is a testament to the patriarchal norms that remain deeply ingrained within the fabric of Ghanaian society and the cultural norms of individuality.[119] Moore avers that 'the law (in the sense of state-enforceable law) is only one of several factors that affect the decisions people make, the actions they take and the relationships they

[114] ibid 4.

[115] *Anthony Victor Obeng v Mrs Theresa Henrietta Obeng* [2015] GHASC 112 (SC Ghana) took the longest to decide, sixteen years in all from trial to the Supreme Court. The petition for divorce in *Mensah v Mensah* (1993) (CA Ghana) (n 96)—which eventually set the precedent on the equality is equity principle—was filed in 1990 and resolved in the Supreme Court eight years later in 1998. The recent case of *Peter Adjei v Margaret Adjei* (n 22), settled in 2021, commenced its journey at the High Court six years earlier in 2015. From the findings, the case with the shortest duration to resolution is *Grace Fynn v Stephen Fynn and Christiana Osei* [2014] GHACA 129 (SC Ghana), which commenced and resolved in four years: 2010–2014.

[116] *White v White* [2001] 1 AC 596 (HL), which set the sharing principle, and *Miller v Miller* [2006] UKHL 24, [2006] 2 AC 618, which clarified the principle years later, were decided in two years and one year, respectively.

[117] Atuguba (n 106).

[118] Dowuona-Hammond (n 23) 102.

[119] Dowuona-Hammond (n 23) 107.

have'.[120] Within the Ghanaian State, women 'belong' to and are regulated by a 'different smaller organised social field' [121], with their own customs and rules that coerce and induce compliance. Rather than legislation, the norms of these social fields form the principal impetus that drives their behaviour, particularly in rural communities.[122] Consequently, the pluralistic nature of the Ghanaian property rights regime increases the arduous nature of this task because any legislation passed must reflect the peculiar situations of different customs of different communities and the different forms of marriage that the Marriages Act recognises.[123]

One such sociocultural norm is polygamy. Polygamy is a creature of African cultural and societal beliefs. In contrast to the CEDAW committee, the Maputo Protocol recognises the ingrained nature of the religious and customary norms surrounding polygamy. Thus, the Protocol does not follow CEDAW's route of requiring all State parties to take 'all legislative and policy measures needed to abolish polygamous marriages.'[124] Instead, the Protocol engenders State parties under Article 6 (c) to enact legislation signalling that monogamy is the encouraged and preferred form of marriage. In Ghana, both the Mohammedan and the customary law marriages, recognised under the Marriages Act, are potentially polygamous. Fourteen per cent of Ghanaian women continue to be in polygamous marriages.[125] In polygamous marriages, more than one woman may simultaneously be entitled to marital property. The issue then becomes how such property will be divided and which woman is entitled to which proportion. Again, the distribution of property presents a significant challenge for the Courts in the case of a divorce within the polygamous unit. To holistically protect women's matrimonial property rights, policies and legislation must consider the women in polygamous marriages. Such policies and legislation must explicitly outline how they must be applied within the context of polygamous marriages. From the Court's approach to protecting the right to own property, it may be inferred that women may be entitled to property purchased when they join the polygamous unit.[126] Nonetheless, it is essential that legislation and policies clearly state how they will be applied within the context of polygamous marriages. The advantages of legislation will be uniform in this way because it will address the intersectionality of factors that

[120] Moore (n 17) 743.
[121] Moore (n 17) 721.
[122] Moore (n 17) 730.
[123] Dowuona-Hammond (n 23) 104.
[124] CEDAW, art 16; CEDAW Committee, 'General Recommendation on Article 16 of the Convention on the Elimination of All Forms of Discrimination against Women' (CEDAW/C/GC/29, 26 February 2013) para 28 <https://www2.ohchr.org/english/bodies/cedaw/docs/comments/CEDAW-C-52-WP-1_en.pdf> accessed 12 December 2022.
[125] Ghana Statistical Service, Ghana Health Service and ICF, 'Ghana 2017 Maternal Health Survey: Key Findings' 4 (2017) <https://dhsprogram.com/pubs/pdf/SR251/SR251.pdf> accessed 12 December 2022.
[126] *Mensah v Mensah* (n 93) 1; *Peter Adjei v Margaret Adjei* (n 22) 11.

combine to discriminate against women's property rights and ensure that the law is equitable for women in all circumstances.

IV. Conclusion

This article has used an African feminist lens to examine the gaps that arise in protecting women's matrimonial property rights when State legislative reform does not reflect African women's peculiar experiences and values. It has done this by first discussing a plethora of national, regional, and international laws and policies that harmonise women's rights and provide a baseline of rights to which women are entitled. The author identifies and examines three key issues that lead to implicit gender biases that seemingly gender-neutral laws on women's matrimonial property rights may reinforce if left unaddressed. These are non-monetary contributions through social reproductive labour, Ghana's slow-paced litigation system, and sociocultural norms like polygamy.

Effective protection of women's rights to matrimonial property requires a holistic approach to law-making and regulation. Achieving this protection will require efforts on multiple fronts—public education to progress norms, deliberate institutional action to stand for commitments made in both domestic and international legal documents and an approach to broad-based legislation that considers multiple actors and contexts focused on Ghanaian women's experiences. Ultimately, political will and broader societal commitment to going beyond tokenism when it comes to equity for women are what will give true meaning to all the legal provisions and doctrines crafted and evolved in recent decades.

Cambridge Law Review (2023) Vol 8, Issue 1, 152–177

Artistic Expression at Risk: The Overlap of Sound Mark Protection and Phonogram Protection in EU IP Law

RÉMI SAIDANE*

ABSTRACT

This article attempts to determine to what extent sound mark protection and phonogram protection overlap at the expense of artistic expression in EU law. It first analyses the sound mark protection regime to show that acquiring sound mark protection is difficult, although the criteria for registration have become more flexible and clarified. However, the rights granted to sound mark proprietors are extensive, which allows them to challenge a variety of sounds at the expense of artistic expression. Secondly, the phonogram protection regime is studied to demonstrate that phonogram protection is easy to acquire thanks to its light requirements. It is then argued that the Court of Justice of the European Union in *Pelham* failed to provide efficient and flexible safeguards to protect artistic expression against phonogram producers' claims. Finally, the detrimental effects on artistic expression caused by the overlap of sound mark protection and phonogram protection are analysed. When sound mark protection is granted, phonogram protection will easily, if not automatically, overlap with it. The accumulation of the two protections, because of their unfavourable approach to artistic expression, leads to detrimental effects on it. Moreover, there is a conflict of individual imperatives between the two systems at the expense of artistic expression. This article thus finds that the combined effects of the expansion of sound mark protection and the unprotective approach to artistic expression regarding phonogram protection in *Pelham* cause serious interferences with artistic expression. Neither of the two regimes adequately protects artistic expression, and the possibility of combining them seriously puts artistic expression at risk. This article suggests that reforms to integrate fair use in both sound mark and phonogram regimes are necessary.

* LLM Candidate, University of Cambridge; LLB (Maastricht).

I. Introduction

It is the summer of 2013. The train from Paris arrives in Aix-en-Provence. As one of the passengers alights, he hears four notes sung in a woman's voice coming from the train station's speakers: 'C, G, A-flat, E-flat'. Mesmerised by this sound, the man steps closer to the loudspeakers: finally, rock icon David Gilmour knows what his next single will sound like. After hearing the jingle for France's national railway operator, the SNCF, the musician decided to use this five-second sound sample as the central element of his song 'Rattle that Lock'.[1] Although he reached an agreement to use this sample with *Sixième Son*, the producer of the jingle, the latter would soon start legal proceedings before French courts based on copyright-related protection of phonograms. This claim would ultimately be rejected.[2] While this dispute between the producers and the artist was based on a copyright-related claim, it could have been that the SNCF had brought a trade mark claim against David Gilmour. The SNCF jingle is broadly perceived by consumers as a tool to identify the company's services,[3] which could be expected to be subject to trade mark protection. This case is one example of how the trade mark protection of sounds (hereinafter 'sound mark protection') and the copyright-related protection of phonograms (hereinafter 'phonogram protection') could overlap at the expense of artistic expression.

Sound marks are trade marks vested in sounds.[4] Their purpose is to protect distinctive sounds of goods or services of an undertaking, to avoid their use by another undertaking in the course of trade.[5] They are 'non-traditional marks', as opposed to traditional marks like words, given that their recognition as trade marks is a recent phenomenon.[6] Meanwhile, phonograms are any exclusively aural fixation of sounds of performances or other sounds.[7] They are subject to copyright-related protection in favour of phonogram producers.[8] Artistic expression is a component of freedom of expression allowing one to take part in

[1] Laura Snapes, 'David Gilmour in Legal Fight with Composer of French SNCF Train Jingle' *The Guardian* (London, 19 November 2019) <www.theguardian.com/music/2019/nov/19/david-gilmour-legal-fight-composer-french-sncf-train-jingle-michael-boumendil> accessed 26 September 2021.

[2] CA Paris 15 June 2021, n° 19/14255.

[3] Sixième Son, 'SNCF' (*sixiemeson.com*) <www.sixiemeson.com/sncf/> accessed 26 September 2021.

[4] Regulation (EU) 2017/1001 of the European Parliament and of the Council of 14 June 2017 on the European Union trade mark [2017] OJ L154/1 ('Trade Mark Regulation'), art 4; Directive (EU) 2015/2436 of the European Parliament and of the Council of 16 December 2015 to approximate the laws of the Member States relating to trade marks [2015] OJ L336/1 ('Trade Mark Directive'), art 3.

[5] Justine Pila and Paul Torremans, *European Intellectual Property Law* (2nd edn, Oxford University Press 2019) 375–376.

[6] Irene Calboli and Martin Senftleben, 'Introduction' in Irene Calboli and Martin Senftleben (eds), *The Protection of Non-Traditional Trademarks* (Oxford University Press 2018) 1–2.

[7] International Convention for the Protection of Performers, Producers of Phonograms and Broadcasting Organizations (adopted 26 October 1961, entered into force 18 May 1964) 496 UNTS 43 ('Rome Convention'), art 3(b).

[8] Paul ACE van der Kooij and Dirk JG Visser, *EU IP Law: A Short Introduction to European Intellectual Property Law* (deLex B.V. 2015) 106–109.

the 'public exchange of cultural, political and social information and ideas of all kinds'.[9]

The continuous expansion of intellectual protection regimes in European Union (EU) intellectual property ('IP') law has led to areas of cumulative protection.[10] In that context, and because they are both vested in sounds, phonogram protection and sound mark protection may cohabit in some instances. The reliance on the two regimes could be a way for undertakings to cumulate their benefits. This phenomenon is, in fact, one facet of a larger one: the copyright/trade mark interface.[11] While the two regimes serve different purposes, they could cohabit the same medium. Such cohabitation could affect artistic expression because the cumulated protections could be powerful tools for the commercial purpose of opposing artistic works in favour of sound marks.

It therefore seems necessary to answer the following question: to what extent do European Union laws on phonogram protection and sound mark protection overlap at the expense of artistic expression?

To answer this question, this article will follow a doctrinal legal research methodology. It will, in its first part, analyse sound mark protection under EU law. For this purpose, the restrictive criteria for sound mark protection, as well as the extensive rights granted by such protection, will be examined. For this purpose, an analysis of the secondary law developed at the EU level and the judgments of the Court of Justice of the European Union (hereinafter 'CJEU' or 'the Court') in connection with phonogram protection will be undertaken (Section II). This will be followed by a study of the easily acquirable protection of phonograms under EU law. The flexible criteria for phonogram protection will be explored. The EU secondary law and CJEU judgments on sound mark protection will be analysed. The extent of the protection granted will be examined as well as the clarifications in relation to artistic expression in the *Pelham* case (Section III). Finally, the overlap of the two regimes and the risks this poses to artistic freedom will be assessed. The two legal frameworks will, following individual analyses of each of them, be confronted to demonstrate how they overlap in a way that could affect artistic expression (Section IV). This research will be limited to EU law, thus excluding national law. Moreover, the only intellectual property rights studied will be phonogram producers' rights and sound mark proprietors' rights. All other rights are not relevant for the purpose of this research. Finally, the friction between the two systems will only be studied in cases where phonogram protection

[9] Christophe Geiger and Elena Izyumenko, 'The Constitutionalization of Intellectual Property Law in the EU and the *Funke Medien*, *Pelham* and *Spiegel Online* Decisions of the CJEU: Progress, But Still Some Way to Go!' (2020) 51 IIC - International Review of Intellectual Property and Competition Law 282, 293. See Case C-476/17 *Pelham GmbH, Moses Pelham, Martin Haas v Ralf Hütter, Florian Schneider-Esleben* EU:C:2019:624, para 34.

[10] Martin Senftleben, *The Copyright / Trademark Interface: How the Expansion of Trademark Protection Is Stifling Cultural Creativity* (Kluwer Law International 2020) 15, 16.

[11] ibid 1–10.

is used by undertakings to protect signs that could qualify as sound marks from artists sampling them. The other aspects of the overlap will not be studied.

II. The Extensive Protection of Sound Marks under EU Law

Sound marks may be protected under EU law if the strict criteria for registration are met (Section II.A). If their sound mark is registered, sound mark proprietors will enjoy a broad set of rights to oppose artistic expression (Section II.B).

A. The Criteria for the Registration of Sound Marks

To be protected under EU law, sound marks can be registered for certain classes of goods or services, either as EU trade marks or as national trade marks.[12] Sound marks are part of the class of trade marks informally called 'non-traditional marks', for they are trade marks created by means of unorthodox media.[13] Originally, authorities showed a fierce reluctance to register sound marks and other non-traditional marks.[14] However, in the EU, the evolution of the legal framework for sound marks and the judgments of the General Court of the CJEU have introduced some flexibility into the regime.

(i) The Need for a Distinctive Sign with the High Threshold of Resonance

For a sound mark to be registered, it must be a sign capable of distinguishing the goods or services of an undertaking from those of other undertakings.[15]

Previous EU legal instruments on trade marks did not explicitly list 'sound' as a type of sign.[16] However, the new Regulation (EU) 2017/1001 on the European Union trade mark (hereinafter 'EUTMR') and Directive (EU) 2015/2436 to approximate the laws of the Member States relating to trade marks (hereinafter 'TMD') have clearly welcomed it into the family of signs subject to trade mark protection.[17]

Turning to distinctiveness, the Court further developed in *Rewe Zentral* that it is generally assessed in relation to: (a) the goods or services for which registration

[12] Trade Mark Regulation, art 6; Trade Mark Directive, art 1.
[13] Calboli and Senftleben (n 6) 1–4.
[14] Dev S Gangjee, 'Paying the Price for Admission: Non-Traditional Marks Across Registration and Enforcement' in Irene Calboli and Martin Senftleben (eds), *The Protection of Non-Traditional Trademarks* (Oxford University Press 2018) 61, 62.
[15] ibid art 4(a); Trade Mark Directive, art 3(a). See Pila and Torremans (n 5) 348.
[16] Council Regulation (EC) 207/2009 of 26 February 2009 on the European Union trade mark [2015] OJ L341/21, art 4; Directive 2008/95/EC of the European Parliament and of the Council of 22 October 2008 to approximate the laws of the Member States relating to trade marks [2008] OJ L299/25, art 2.
[17] Trade Mark Regulation, art 4; Trade Mark Directive, art 3.

is sought; and (b) the perception of the targeted public composed of consumers of the goods or services.[18] A minimum degree of distinctiveness is, however, sufficient for registration.[19] Nevertheless, the Court in *Glaverbel* added that, although the criteria to assess distinctiveness were the same for all trade marks, the perception of the target consumers could differ for certain classes of trade mark.[20] The Court was implicitly referring to non-traditional marks, as evidenced by the previous judgments it cited.[21]

For sound marks, the application of the criterion of the 'perception of the target consumers' was thus further elaborated in *Globo Comunicação*.[22] The Court acknowledged the increasing role of sounds as a means of identifying goods or services, especially in the media sector.[23] Regarding the perception of the target consumers, the sound needed to have a 'certain resonance' enabling them to perceive and regard the sound as a trade mark having the ability to identify a good or service, rather than a mere functional element or an element devoid of inherent characteristic.[24] Excessively simple and banal sounds, such as the repetition of two identical notes, would make such a resonance impossible. It would not enable the target consumer to perceive it as identifying the good or service, because it would merely refer to itself and would not lead to a certain form of attention enabling the target consumer to perceive the sound's identifying function.[25] Nevertheless, the sound did not have to be original or fanciful to meet the threshold.[26] The CJEU later confirmed the standard of resonance when applying the criterion of the perception of the target consumers in *Ardagh Metal Beverage*.[27] 'Resonance' was the standard for sound marks, and the other standards developed for other types of marks could not be applied to sound marks.

Criticism has been raised regarding the concept of 'resonance'. Indeed, this word was not defined by the Court. While it is defined in the *Cambridge Dictionary* as a 'feeling, thought, memory' that something makes someone have, or the

[18] Case T-79/00 *Rewe Zentral AG v Office for Harmonisation in the Internal Market (Trade Marks and Designs) (OHIM)* EU:T:2002:42, [2002] ECR II-705, para 27.

[19] ibid para 28.

[20] Case C-445/02 P *Glaverbel SA v Office for Harmonisation in the Internal Market (Trade Marks and Designs) (OHIM)* EU:C:2004:393, [2004] ECR I-6267, para 23.

[21] ibid. On non-traditional trade marks, see Joined Cases C-456/01 P and C-457/01 P *Henkel KGaA v Office for Harmonisation in the Internal Market (Trade Marks and Designs) (OHIM)* EU:C:2004:258, [2004] ECR I-5089, para 38. See also Joined Cases C-468/01 P to C-472/01 P *Procter & Gamble Company v Office for Harmonisation in the Internal Market (Trade Marks and Designs) (OHIM)* EU:C:2004:259, [2004] ECR I-5154, para 36.

[22] Case T-408/15 *Globo Comunicação e Participações S/A v European Union Intellectual Property Office* EU:T:2016:468, paras 41–46.

[23] ibid paras 42–44.

[24] ibid para 45. See Guy Heath and others, 'When Should a Sound Mark Be Regarded as Excessively Simple, Notwithstanding the Familiarity of Consumers with the Use of Sound Marks in the Economic Sector Concerned? (2016) 107 The Trademark Reporter 505, 506.

[25] *Globo Comunicação* (n 22) para 46. See Heath and others (n 24) 506.

[26] ibid para 57. See Heath and others (n 24) 506.

[27] Case T-668/19 *Ardagh Metal Beverage Holdings GmbH & Co KG v European Union Intellectual Property Office* EU:T:2021:420, para 24.

'quality' in a thing that 'makes this happen',[28] no clear legal definition was given by the Court. Even when relying on the definition from the *Cambridge Dictionary*, it implies a subjective assessment relying on feelings that is likely to vary among the average consumers of the good or service concerned by the sound mark.[29] Moreover, the words used in the different language versions of the judgment to describe this concept make it even more unclear. In French, the word used, *prégnance*, can be translated as 'something that imposes itself on the mind, which produces a strong impression'.[30] In Spanish, the word *fuerza* was employed, which means strength.[31] These words, while tending to describe a similar concept, diverge substantially in terms of definition, which blurs the exact meaning of the concept of 'resonance'. Finally, while EU law has in principle accepted the registration of sound marks, the two judgments of the CJEU led to the rejection of the sound marks at stake for being too banal, and a substantial number of decisions of the Boards of Appeal of the EUIPO rejected sound marks.[32] This shows how difficult it is in fact to register a sound mark because of the criterion of distinctiveness.[33] On a side note, trade marks can be registered when they lack inherent distinctiveness thanks to their distinctiveness acquired through use.[34] This possibility has been acknowledged for sound marks by the CJEU.[35]

(ii) The Flexibility Introduced for the Criterion of Appropriate Representation

Finally, the third criterion is a procedural one, concerning the representation of the sign on the register.[36] The sign must satisfy the so-called *Sieckmann* criteria. Accordingly, for a sign to be deemed a trade mark, it must be represented in a manner that is clear, precise, self-contained, easily accessible, intelligible, durable and objective.[37] The *Sieckmann* criteria would later be incorporated into the EUTMR and the TMD.[38] These criteria play an important role for non-traditional marks, because of the unconventional ways of representing

[28] 'Resonance' (*Cambridge Dictionary*, 2022) <https://dictionary.cambridge.org/fr/dictionnaire/anglais/resonance> accessed 3 April 2022.

[29] Stefan Martin, 'The Hiss Produced by the Opening of a Drink Can Cannot Be Registered as EUTM' (2021) 16 Journal of Intellectual Property Law & Practice 1164, 1166; Gordon Humphreys, Nedim Malovic, and Stefan Martin, 'Round-Up of Non-Traditional EU Trade Mark Decisions in 2021' (2021) 17 Journal of Intellectual Property Law & Practice 350, 356.

[30] 'Prégnance' (*Larousse*, 2022) <www.larousse.fr/dictionnaires/francais/pr%C3%A9gnance/63496#:~:text=Litt%C3%A9raire.,ou%20moins%20grande%20aux%20sujets.> accessed 3 April 2022; 'Prégnant' (*Larousse*, 2022) <www.larousse.fr/dictionnaires/francais/pr%C3%A9gnant/63499> accessed 3 April 2022.

[31] Martin (n 29) 1166.

[32] Humphreys, Malovic, and Martin (n 29) 356.

[33] Martin (n 29) 1166.

[34] Trade Mark Regulation, art 7(3).

[35] *Globo Comunicação* (n 22) para 51.

[36] Pila and Torremans (n 5) 350.

[37] Case C-273/00 *Ralf Sieckmann v Deutsches Patent- und Markenamt* EU:C:2002:748, [2002] ECR I-11737, para 55.

[38] Trade Mark Regulation, recital 10, art 4(b); Trade Mark Directive, recital 13, art 3(2). See Gangjee (n 14) 67.

them.[39] Each class of non-traditional marks is subject to its own representation requirements.[40] Originally, the Court admitted in *Shield Mark BV* that the *Sieckmann* criteria could not be met for sound marks when the sign is represented by a description using the written language, such as an indication of the musical notes or a sequence of musical notes, or the cry of a cockerel, or a mere onomatopoeia.[41] A sign could, however, be represented by employing musical notation, with a stave divided into measures, a clef, musical notes and rests indicating the relative value and accidentals when deemed necessary.[42] In addition, the previous EUTMR and TMD explicitly required a graphic representation,[43] which would exclude several means of representation, such as sound files. This made the representation of sound marks technically difficult. However, with the adoption of the new EUTMR and TMD, the criterion of graphic representation has been removed.[44] Indeed, as the European Commission had highlighted, alternative representations such as sound files could be preferable to graphic representation in some instances.[45]

As a result, the inflexible criterion of graphic representation has been replaced by a standard of 'appropriate representation', or so-called 'adequate representation', according to which any appropriate form of representation using generally available technology can be employed to represent trade marks.[46] Today, sound marks can be represented either by an accurate representation of the sound in musical notation or an audio file reproducing the sound.[47]

All in all, it is possible to register sound marks, thanks to the explicit introduction of sound as signs and the flexibility introduced for representation in EU law. The strict criterion of distinctiveness and its standard of resonance nevertheless render such registration unclear and complicated. If one manages to meet such a standard, their sound mark would enjoy extensive protection (Section II.B).

[39] Gangjee (n 14) 66.
[40] ibid 67.
[41] Case C-283/01 *Shield Mark BV v Joost Kist h.o.d.n. Memex* EU:C:2003:641, [2003] ECR I-14313, para 64. See Pila and Torremans (n 5) 352–353.
[42] *Shield Mark* (n 41) para 64.
[43] Council Regulation (EC) 207/2009 (n 16) art 4; Directive 2008/95/EC (n 16) art 2.
[44] Trade Mark Regulation, art 4; Trade Mark Directive, art 3.
[45] Commission, 'Proposal for a Regulation of the European Parliament and of the Council amending Council Regulation (EC) No 207/2009 on the Community trade mark' COM (2013) 151 final, 7. See also Max Planck Institute for Intellectual Property and Competition Law, *Study on the Overall Functioning of the Trade Mark System in Europe* (Publication Office of the European Union 2011) 67.
[46] Trade Mark Regulation, recital 10. See Gangjee (n 14) 65.
[47] Commission Implementing Regulation (EU) 2018/626 of 5 March 2018 laying down detailed rules for implementing certain provisions of Regulation (EU) 2017/1001 of the European Parliament and of the Council on the European Union trade mark, and repealing Implementing Regulation (EU) 2017/1431 [2018] OJ L104/37, art 3(3)(g).

B. The Extensive Rights of Sound Mark Proprietors to Oppose Artistic Expression

Once a sound mark is registered, it enjoys the broad protection granted under EU law for trade marks. The rights of sound mark proprietors are registered for ten years,[48] and can be renewed every ten years.[49] These rights are subject to a limited number of exceptions and limitations, under which the relativisations regarding artistic expression are limited.

(i) The Broad Set of Rights Granted to Sound Mark Holders

Three grounds are available to raise a claim for trade mark infringement: double identity,[50] likelihood of confusion,[51] and trade mark reputation.[52]

Under a claim for double identity, trade mark holders can prevent a third party from using in the course of trade and without their consent a sign identical to the trade mark and used in relation to goods or services which are identical to those for which the trade mark is registered.[53]

As regards the likelihood of confusion, trade mark holders can prevent a third party from using in the course of trade and without their consent: (a) a sign identical or similar to the trade mark; (b) in relation to goods or services identical or similar to the goods or services for which the trade mark is registered; and (c) where there exists a likelihood of confusion for the public between the trade mark and the sign, which includes the likelihood of association between the sign and the trade mark.[54]

Finally, a trade mark reputation claim is available for trade mark holders to prevent third parties from using, under certain conditions, certain signs in the course of trade and without their consent.[55] First, there must be a sign which is identical or similar to the trade mark. Second, the sign must be used in relation to any good or service, either identical, similar, or dissimilar to that of the earlier mark. Third, the earlier trade mark must have a reputation. The concept of reputation implies a certain degree of knowledge among the relevant public.[56] The relevant public depends on the marketed good or service and can be either the general public or a specialised public.[57] Fourth, the use of that sign must either

[48] Commission Implementing Regulation (EU) 2018/626 (n 47) art 52; Trade Mark Directive, art 48.
[49] Commission Implementing Regulation (EU) 2018/626 (n 47) art 53; Trade Mark Directive, art 49.
[50] Trade Mark Regulation, art 9(2)(a); Trade Mark Directive, art, 10(2)(a).
[51] Trade Mark Regulation, art 9(2)(b); Trade Mark Directive, art 10(2)(b).
[52] Trade Mark Regulation, art 9(2)(c); Trade Mark Directive, art 10(2)(c).
[53] Trade Mark Regulation, art 9(2)(a); Trade Mark Directive, art 10(2)(a).
[54] Trade Mark Regulation, art 9(2)(b); Trade Mark Directive, art 10(2)(b).
[55] Trade Mark Regulation, art 9(2)(c); Trade Mark Directive, art 10(2)(c).
[56] Case C-301/07 *PAGO International GmbH v Tirolmilch registrierte Genossenschaft mbH* EU:C:2009:611, [2009] ECR I-9429, para 21.
[57] ibid para 22. See Case C-375/97 *General Motors Corporation v Yplon SA* EU:C:1999:408, [1999] ECR I-5421, para 24.

take unfair advantage of, or be detrimental to, the distinctive character or repute of the mark. For trade mark reputation claims, no likelihood of confusion is required, which means that a mere link in the mind of the average consumer between the sign and the mark is sufficient.[58]

Sound mark proprietors can therefore rely on claims for double identity, likelihood of confusion, and trade mark reputation to protect their sound marks from being infringed by third parties, including artists. This extensive set of rights allows them to prevent third parties from using identical, similar, or dissimilar sounds in a broad range of scenarios, with few criteria to be met.

(ii) The Few Limitations and Defences that Protect Artistic Expression

There are only a few limitations and defences which artists could rely on against sound mark claims, and the relativisations of trade mark proprietors' rights in relation to artistic expression are limited.

First, there are internal limitations in the trade mark system that can be relied on to limit the scope of sound mark protection. This is the case with the criterion of the infringing sign being used 'in the course of trade'.[59] Indeed, trade mark protection is, in principle, only relevant against signs used in the course of trade,[60] that is in the context of commercial activity with a view to an economic advantage and not as a private matter,[61] and in the user's own commercial communication.[62] This criterion supposedly excludes purely artistic uses of a trade mark from the scope of trade mark proprietors' claims,[63] but can be insufficient to exclude mixed commercial uses. These are non-commercial uses of a trade mark in a commercial context in relation to goods or services,[64] such as the use of a sampled sound mark in a song that would later be commercialised. While the sound mark is not used to distinguish the song as a product, it would end up being incorporated into it when commercialised.

Another internal limitation is the requirement that the infringing sign is used 'in relation to goods or services'.[65] The CJEU has adopted a flexible approach to that criterion, according to which a 'reference', 'link' or 'association' to the goods

[58] Case C-408/01 *Adidas-Salomon AG and Adidas Benelux BV v Fitnessworld Trading Ltd* EU:C:2003:582, [2003] ECR I-12537, para 31.

[59] Łukasz Zelechowski, 'Invoking Freedom of Expression and Freedom of Competition in Trade Mark Infringement Disputes: Legal Mechanisms for Striking a Balance' (2018) 19 ERA Forum 115, 118.

[60] Trade Mark Regulation, art 9(2); Trade Mark Directive, art 10(2).

[61] Case C-206/01 *Arsenal Football Club plc v Matthew Reed* EU:C:2002:651, [2002] ECR I-102773, para 40.

[62] Joined cases C-236/08 to C-238/08 *Google France SARL and Google Inc v Louis Vuitton Malletier SA, Google France SARL v Viaticum SA and Luteciel SARL and Google France SARL v Centre national de recherche en relations humaines (CNRRH) SARL and Others* EU:C:2010:159, [2010] ECR I-2417, para 56.

[63] Zelechowski (n 59) 118.

[64] ibid 119; Martin Senftleben, 'Robustness Check: Evaluating and Strengthening Artistic Use Defences in EU Trademark Law' (2022) 53 International Review of Intellectual Property and Competition Law 567, 572.

[65] Trade Mark Regulation, art 9(2); Trade Mark Directive, art 10(2).

or services is sufficient to pass the threshold.[66] This diminishes the limiting effect of the criterion in protecting artistic expression. More generally, the realm of trade mark protection expanded into the natural realm of copyright and rights related to copyrights,[67] with no parallel development of users' rights.[68] As a result, the gatekeeping role of the two internal limitations in restricting the extent of trade mark claims has been eroded, with their contours being blurred and trade marks being over-protected.[69] An illustration of this phenomenon is the extension of the list of particularly relevant infringements in the new EUTMR and the TMD,[70] which entrenches the criterion of use in relation to goods or services.[71]

Next to these internal limitations, only a limited set of four defences are explicitly provided in the EU trade mark system,[72] with no defence specifically designed for artistic expression.[73] First, third parties who are natural persons can use their own name or address, regardless of any trade mark protection.[74] Second, under the descriptive use defence, third parties can freely use signs or indications concerning mere characteristics of the goods or services.[75] This exception could allow the use of elements of cultural significance registered as trade marks thanks to their acquired distinctiveness.[76] Its protective effect on artistic expression is, however, limited. Purely decorative uses,[77] as well as faithful reproductions of a trade mark,[78] are excluded from the scope of this defence.[79] Next, under the non-distinctive use defence, third parties can freely use non-distinctive signs or indications.[80] Finally, under the referential use defence, third parties can use a trade mark to identify or refer to goods or services as those of the proprietor of

[66] Case C-63/97 *Bayerische Motorenwerke AG and BMW Nederland BV v Ronald Karel Deenik* EU:C:1999:82, [1999] ECR I-905, para 42; Case C-533/06 *O2 Holdings Limited and O2 (UK) Limited v Hutchison 3G UK Limited* EU:C:2008:339, [2008] ECR I-4231, paras 35–36; Case C-17/06 *Céline SARL v Céline SA* EU:C:2007:497, [2007] ECR I-7041, para 23; Joined cases C-236/08 to C-238/08 *Google France SARL and Google Inc v Louis Vuitton Malletier SA, Google France SARL v Viaticum SA and Luteciel SARL, and Google France SARL v Centre national de recherche en relations humaines (CNRRH) SARL and Others* EU:C:2010:159, [2010] ECR I-2417, para 72; Case C-324/09 *L'Oréal SA and Others v eBay International AG and Others* EU:C:2011:474, [2011] ECR I-6011, para 93. See Senftleben, 'Robustness Check' (n 64) 577–578.

[67] Senftleben, *The Copyright / Trademark Interface* (n 10) 134, 141.

[68] Jens Schovsbo, 'Mark My Words - Trademarks and Fundamental Rights in the EU' (2018) 8 UC Irvine Law Review 555, 557, 564–566.

[69] ibid 557, 564–566.

[70] Trade Mark Regulation, art 9(3); Trade Mark Directive, art 10(3).

[71] Schovsbo (n 68) 564–565.

[72] Trade Mark Regulation, art 14(1); Trade Mark Directive, art 14(1).

[73] Senftleben, 'Robustness Check' (n 64) 599.

[74] Trade Mark Regulation, art 14(1)(a); Trade Mark Directive, art 14(1)(a). See Senftleben, 'Robustness Check' (n 64) 581.

[75] ibid art 14(1)(b); Trade Mark Directive, art 14(1)(b). See Senftleben, 'Robustness Check' (n 63) 583.

[76] Senftleben, 'Robustness Check' (n 64) 583.

[77] Case C-102/07 *Adidas AG and Adidas Benelux BV v Marca Mode CV and Others* EU:C:2008:217, [2008] ECR I-2439, para 48.

[78] Case C-48/05 *Adam Opel v Autec AG* EU:C:2007:55, [2007] ECR I-1017, para 44.

[79] Senftleben, 'Robustness Check' (n 64) 583–584.

[80] Trade Mark Regulation, art 14(1)(b); Trade Mark Directive, art 14(1)(b). Senftleben, 'Robustness Check' (n 64) 583.

that trade mark.[81] This exception could be relevant in cases of mixed commercial uses deemed to be in the course of trade when there is an artistic, polemical, parodic, or satirical purpose.[82] In practice, however, there is little evidence of the application of this exception in these contexts.[83]

It has been argued that the four defences should be construed as balancing tools to be generally used in the trade mark system. Intellectual property rights cannot be treated as pure economic assets as tangible objects would be under property law, because of the societal need to access knowledge.[84] The interface between the exclusive rights and the socio-economic need for free access to intellectual property objects should be considered when applying the defences, thus justifying a liberal and user-friendly approach.[85] But even with this approach, the defences are, in any case, subject to the burden of proving that the use of the sign is in accordance with honest practices in industrial or commercial matters.[86] The new EUTMR and TMD especially stress that, in the context of artistic expression, while trade mark law must be applied in a way that ensures full respect for fundamental rights and freedoms, including freedom of expression, the use of a trade mark can only be deemed fair if it is in accordance with honest practices in industrial and commercial matters.[87] This requirement is problematic from a theoretical point of view given that it is uncertain whether artistic productions should be subject to industrial and commercial standards.[88] By obliging artists to abide by such standards, their creative autonomy could be affected,[89] thus interfering with artistic expression. Moreover, from a practical point of view, artists are not always acquainted with commercial and industrial behavioural standards, and it could be difficult to expect them to fully master these standards.[90]

Finally, there is a limitation internal to trade mark reputation claims, namely the 'due cause' limitation. According to this limitation, a trade mark with a reputation cannot be protected against the use of a sign with due cause.[91] While the CJEU has not yet ruled on due cause in the context of artistic expression, it has shown some flexibility in the application of this criterion in the context of freedom of expression and freedom of competition.[92] The Court recognises objectively overriding reasons as well as subjective interests of third parties as

[81] Trade Mark Regulation, art 14(1)(c); Trade Mark Directive, art 14(1)(c).
[82] Zelechowski (n 59) 127; Senftleben, 'Robustness Check' (n 64) 581.
[83] Senftleben, 'Robustness Check' (n 64) 581.
[84] Zelechowski (n 59) 128.
[85] ibid.
[86] Trade Mark Regulation, art 14(2); Trade Mark Directive, art 14(2).
[87] Trade Mark Regulation, recital 21; Trade Mark Directive, recital 27.
[88] Senftleben, 'Robustness Check' (n 64) 586–587.
[89] ibid 586.
[90] ibid.
[91] Trade Mark Regulation, art 9(2)(c); Trade Mark Directive, art 10(2)(c).
[92] Case C-323/09 *Interflora Inc and Interflora British Unit v Marks & Spencer plc and Flowers Direct Online Ltd* EU:C:2011:604, [2011] ECR I-8625, para 91; Zelechowski (n 59) 129.

forming part of the concept of due cause.[93] It tends to treat due cause as an open-ended general clause to balance the competing interests of trade mark proprietors and users,[94] which could be relied on in an artistic context.

To sum up, artists only have little room to answer to sound mark proprietor claims. The internal limitations tend to show deficiencies with the recent expansion of trade mark protection. The few defences available were not designed to protect artistic expression. Even when these defences are construed to protect artistic expression, the burdensome requirement of honest practices is a bar to the reliance on these defences for artists. The due cause limitation may be relied on in the context of artistic expression, but it would only be available against sound mark reputation claims.

As a result, it can be said that sound mark protection is difficult to acquire but offers extensive rights to its proprietors, at the expense of artistic expression. Next to it, the phonogram regime could be an unorthodox but easy way to protect sounds used in the course of trade in relation to goods or services (Section III).

III. The Easily Acquirable Protection of Phonograms under EU Law

Phonograms are protected by so-called related rights. The light criteria for their protection are defined in EU legislation and international instruments (Section III.A). The rights granted under these regimes are set out in EU Directives and have been subject to some clarifications by the CJEU in the *Pelham* case (Section III.B).[95]

A. The Light Criteria for the Protection of Phonograms

The criteria to determine whether there is a phonogram protected by a right related to copyright diverge to some extent from the criteria for copyright protection of works. These criteria can be inferred from the Directive 2001/29/EC on the harmonisation of certain aspects of copyright and related rights in the information society (hereinafter the InfoSoc Directive),[96] and, failing that, from international treaties.[97]

First, there should be a phonogram. While the InfoSoc Directive does not explicitly define what a phonogram is, it can be determined that these criteria can

[93] Case C-65/12 *Leidseplein Beheer BV and Hendrikus de Vries v Red Bull GmbH and Red Bull Nederland BV* EU:C:2014:49, paras 45–48. See Zelechowski (n 59) 129.
[94] ibid paras 27–49. See Zelechowski (n 59) 130.
[95] *Pelham* (n 9).
[96] Copyright Directive.
[97] Pila and Torremans (n 5) 267.

be found in the WIPO Performance and Phonograms Treaty (hereinafter WPPT)[98] which the EU and a majority of Member States have already signed.[99] Article 2(b) WPPT defines phonograms as the fixation of the sounds of a performance or other sounds, or of a representation of sounds, other than in the form of a fixation incorporated in a cinematographic or other audio-visual work.[100] Contrary to copyright protection,[101] no requirement of originality has to be fulfilled for phonogram protection.[102] Indeed, it is vested in the mere fixation of the sounds, rather than in the arrangement of these sounds.[103] This renders the threshold for phonogram protection lower than for copyright protection of works.[104]

Another condition is that, contrary to sound mark protection, there is generally no formality required. National law may provide some formalities, but their strictness is limited to a notice including the year of the first publication and the symbol (P) placed on the copies.[105]

Thus, in the EU, the criteria for phonogram protection are light and flexible. No creativity is necessary. While formalities could be imposed, these are, in practice, easy to overcome and require a mere inscription on commercialised copies, and no registration is required. This framework would allow undertakings to easily protect sounds used in the course of trade in relation to goods or services against artistic expression. Once the criteria for protection are fulfilled, phonogram producers can access the rights granted in relation to the said phonogram (Section III.B).

B. The Rights of Phonogram Producers in Relation to Artistic Expression in the *Pelham* Case

Under EU law, the set of rights granted to phonogram producers regarding their phonograms is part of the general copyright and related rights system. First, under article 2(c) of InfoSoc Directive, phonogram producers have the exclusive right to authorise or prohibit direct or indirect, temporary or permanent reproduction by any means and in any form, in whole or in part, of

[98] WIPO Performances and Phonograms Treaty (adopted 20 December 1996, entered into force 20 May 2002) TRT/WPPT/001 ('WPPT').

[99] Copyright Directive, recitals 15 and 19.

[100] WPPT, art 2(b).

[101] Case C-145/10 *Eva-Maria Painer v Standard VerlagsGmbH, Axel Springer AG, Süddeutsche Zeitung GmbH, Spiegel-Verlag Rudolf Augstein GmbH & Co KG, Verlag M DuMont Schauberg Expedition der Kölnischen Zeitung GmbH & Co KG* EU:C:2011:798, para 87; Case C-310/17 *Levola Hengelo BV v Smilde Foods BV* EU:C:2018:899, para 36.

[102] Case C-476/17 *Pelham GmbH, Moses Pelham, Martin Haas v Ralf Hütter, Florian Schneider-Esleben* EU:C:2018:1002, Opinion of AG Szpunar, para 30.

[103] ibid para 30.

[104] Lionel Bently and others, 'Sound Sampling, a Permitted Use Under EU Copyright Law? Opinion of the European Copyright Society in Relation to the Pending Reference before the CJEU in Case C-476/17, *Pelham GmbH v. Hütter*' (2019) 50 International Review of Intellectual Property and Competition Law 467, 469.

[105] Rome Convention, art 11. See Pila and Torremans (n 5) 272.

their phonograms.[106] Moreover, under article 3(b) of the InfoSoc Directive, phonogram producers enjoy an exclusive right to authorise or prohibit the making available to the public, of their phonograms in a way that the public may access them from a place and at a time individually chosen by them.[107] These rights are subject to exceptions and limitations set out in article 5 of the InfoSoc Directive, in special cases not conflicting with a normal exploitation of the phonogram and that do not unreasonably prejudice the legitimate interests of phonogram producers.[108] In addition, phonogram producers enjoy exclusive rights to allow or prohibit the rental and lending of their phonograms.[109] They also have a distribution right, which is the exclusive right to make available to the public, by sale or otherwise, their phonograms and copies of their phonograms.[110] The rights of phonogram producers expire seventy years after the lawful publication of the phonogram. Failing that, they expire seventy years after their lawful communication to the public. Failing that, they expire fifty years after their fixation.[111]

The CJEU has recently clarified the extent of the rights of phonogram producers in relation to artistic expression. Indeed, on 29 July 2019, the Court released three major rulings on copyright protection and copyright-related protection: *Funke Medien*,[112] *Spiegel Online*,[113] and *Pelham*.[114] The latter judgment was especially relevant for phonogram protection and artistic expression. Through its answers, it established the extent of phonogram protection in relation to artistic expression.

(i) The Rights of Phonogram Producers Extending to All Samples Recognisable to the Ear

Under its interpretation of article 2(c) of the InfoSoc Directive, the CJEU infers that phonogram producers' reproduction right allows them to prevent anybody from extracting a sound sample from their phonogram, even if very short, to include that sample in another phonogram in a form recognisable to the ear.[115]

To reach such a conclusion, the Court relied on the concept of a 'fair balance'. This concept was already well-established in the CJEU's judgments on

[106] Copyright Directive, art 2(c).
[107] ibid art 3(2)(b).
[108] ibid art 5.
[109] Directive 2006/115/EC of the European Parliament and of the Council of 12 December 2006 on rental right and lending right and on certain rights related to copyright in the field of intellectual property [2006] OJ L376/28, art 3(1)(b).
[110] ibid art 9(1)(b).
[111] Directive 2006/116/EC of the European Parliament and of the Council of 12 December 2006 on the term of protection of copyright and certain related rights [2006] OJ L372/12, art 3(2).
[112] Case C-469/17 *Funke Medien NRW GmbH v Bundesrepublik Deutschland* EU:C:2019:623.
[113] Case C-516/17 *Spiegel Online GmbH v Volker Beck* EU:C:2019:625.
[114] *Pelham* (n 9).
[115] ibid para 39.

copyright law, such as *Promusicae*[116] or, more recently, *Renckhoff*.[117] Accordingly, a fair balance must be struck between the protection of the rights of phonogram producers and the protection of the fundamental rights of phonogram users and the public interest.[118] The CJEU thus proceeds to balance these interests. It first admits that, as a matter of fairness, reproduction by a user of a sound sample extracted from a phonogram, even if very short, must in principle be regarded as a reproduction 'in part' of that phonogram within the ordinary meaning of the provision.[119] Such a reproduction shall therefore be subject to the exclusive right granted to the producer of the phonogram under that provision.[120] Nevertheless, to correctly strike the balance, the CJEU states that, if the sound is modified in a form unrecognisable to the ear in the new phonogram, it cannot be considered a reproduction under article 2(c).[121] It justifies this choice by first highlighting that the protection of IP under the Charter is not absolute.[122] Moreover, as the ECtHR highlighted, it should be possible to take part in the public exchange of cultural, political, and social information and ideas of all kinds.[123] Therefore, to balance adequately the two interests, the CJEU concludes that a sample taken from a phonogram and used in a new phonogram in a modified form unrecognisable to the ear for a distinct artistic creation is not a 'reproduction' under article 2(c) of the InfoSoc Directive.[124]

Thus, in the light of fundamental rights, the CJEU has limited the extent of phonogram protection to cases where samples recognisable to the ear are used. This conclusion has many implications. First, the concept of 'unrecognisable to the ear' has raised concerns because of its lack of elaboration. Indeed, more precise information on the standard of the person's hearing are necessary.[125] Are they the average layperson with little to no expertise in music? Or are they music experts who are acquainted with the industry, such as composers?[126] Moreover, it is debatable whether limiting the interest of phonogram producers to sounds that are only recognisable to the ear is enough to protect artistic expression. It may require more room to be reasonably safeguarded. Allowing phonogram producers to have claims against any sample recognisable to the ear would become

[116] Case C-275/06 *Productores de Música de España (Promusicae) v Telefónica de España SAU* EU:C:2008:54, [2008] ECR I-271, paras 61–70. See Caterina Sganga, 'A Decade of Fair Balance Doctrine, and How to Fix It: Copyright versus Fundamental Rights before the CJEU from *Promusicae* to *Funke Medien*, *Pelham* and *Spiegel Online*' (2019) 41 European Intellectual Property Review 672, 675.
[117] Case C-161/17 *Land Nordrhein-Westfalen v Dirk Renckhoff* EU:C:2018:634, paras 41–47.
[118] *Pelham* (n 9) para 32.
[119] ibid para 28.
[120] ibid para 29.
[121] ibid para 30.
[122] ibid para 33.
[123] ibid para 34.
[124] ibid paras 37, 39.
[125] Bernd Justin Jütte, 'CJEU Permits Sampling of Phonograms Under a De Minimis Rule and the Quotation Exception' (2019) 14 Journal of Intellectual Property Law & Practice 827, 828.
[126] Martin Senftleben, 'Flexibility Grave — Partial Reproduction Focus and Closed System Fetishism in CJEU, *Pelham*' (2020) 51 International Review of Intellectual Property and Competition Law 751, 757.

inconvenient in the music industry, where sound sampling is generalised. With fifty-four per cent of the new albums on *Billboard's Top 25 charts* of 2021 containing samples,[127] this technique has become one of the main tools used by composers today. Hindering its use could be detrimental to phonogram producers themselves in the longer term when their own phonograms would contain samples recognisable to the ear.

(ii) The Impossibility of Relying on External Exceptions Based on Freedom of Expression to Protect Artistic Expression in Relation to Sound Sampling

A second contribution by the CJEU is its refusal to rely on external exceptions based on fundamental rights to protect artistic expression. In *Pelham*, the issue was whether the exceptions to phonogram producers' rights granted under article 5 of the InfoSoc Directive were exhaustive and sufficiently protected the interests and fundamental rights of users of protected subject matter as well as those of the public interest.[128] If so, this would prevent the Member State from establishing its own 'external' exceptions under its national law. The CJEU found that, indeed, the exceptions under article 5 were exhaustive and provided enough protection for the interests of users of protected subject matter as well as the public interest.[129] This decision was, according to the Court, consistent with its previous judgments, such as *Renckhoff*.[130]

To justify such a choice, the CJEU relied again on the concept of fair balance. According to the Court, the exclusive rights of phonogram producers under articles 2 to 4 InfoSoc Directive, on the one hand, and the exceptions as codified in article 5 InfoSoc Directive, on the other hand, were sufficient to strike the balance adequately.[131] This internal system for balancing was sufficient to protect the fundamental rights at stake and consequently excluded the need for a system external to the Directive to protect such rights.[132] In addition, the Court found that the internal limitations further ensure an adequate internal balance because, as article 5(5) required, the internal limitations themselves were only applicable in special cases not conflicting with the normal exploitation of the work that would not unreasonably prejudice the legitimate interests of the right holder.[133] As Recital 32 dictated, these internal limitations also had to be applied

[127] Tracklib, 'State of Sampling' (*Tracklib.com*, 29 November 2021) <https://www.tracklib.com/blog/tracklib-presents-state-of-sampling-2021> accessed 15 January 2022.
[128] *Pelham* (n 9) para 56.
[129] ibid para 65.
[130] ibid para 58. See Case C-161/17 *Land Nordrhein-Westfalen v Dirk Renckhoff* EU:C:2018:634, para 16.
[131] ibid para 60.
[132] ibid para 65. See Thom Snijders and Stijn van Deursen, 'The Road Not Taken – the CJEU Sheds Light on the Role of Fundamental Rights in the European Copyright Framework – a Case Note on the *Pelham*, *Spiegel Online* and *Funke Medien* Decisions' 50 International Review of Intellectual Property and Competition Law 1176, 1183.
[133] ibid para 62. See Copyright Directive, art 5(5).

consistently,[134] which implied that Member States could not provide their own external limitations in national law, given that it would disrupt the system.[135] More importantly, the Court rejected the possibility of an external system on the ground that it would bar the effective harmonisation of the copyright-related protection of phonograms.[136] Such a choice is not surprising, because the CJEU has consistently rejected traditional national doctrines in favour of the harmonisation of rights and limitations in other fields of intellectual property law, such as trade mark law.[137]

Thus, the CJEU refutes that Member States can rely on fundamental freedoms to set out external limitations to phonogram protection. Instead, the CJEU uses freedom of expression as a tool to 'shape' the internal limits of phonogram protection.[138] While showing no flexibility regarding possible external limitations, the CJEU allows a liberal interpretation of the phonogram protection system to ensure sufficient protection of freedom of expression and freedom of the arts.[139] In this way, and with its liberal interpretation relying on internal balancing, the Court ensures that the practical result is similar to using freedom of expression as an external factor.[140]

Many challenges arise from the Court's judgment on that point. First, it is doubtful whether the use of internal factors by the CJEU strikes an adequate balance between the two interests and the fundamental rights related to them to protect artistic expression. Indeed, in the case at hand, the Court did use internal factors in a way that would not tolerate a typical sound sampling situation, thus ruling at the disadvantage of artistic expression.[141] Moreover, by refusing all forms of external factors and relying on the mere internal system found in legislation, the CJEU has rendered the system even less flexible.[142] As the Advocate General has observed in his Opinion in *Funke Medien*, obvious systemic shortcomings regarding the protection of fundamental rights in relation to copyright and related rights had already emerged from the current framework, which raised the need for a legislative amendment.[143] To better this situation, scholars have advocated for the introduction of an open-ended exception in article 5 of the InfoSoc

[134] Copyright Directive, recital 32.
[135] *Pelham* (n 9) para 64.
[136] ibid para 63.
[137] Case C-661/11 *Martin Y Paz Diffusion SA v David Depuydt, Fabriek van Maroquinerie Gauquie NV* EU:C:2013:577, paras 54–55. See Senftleben, 'Flexibility Grave' (n 126) 762. See also Martin Senftleben, 'Trademark Transactions in EU Law – Refining the Approach to Selective Distribution Networks and National Unfair Competition Law' in Irene Calboli and Jacques de Werra (eds), *The Law and Practice of Trademark Transactions – A Global and Local Outlook* (Edward Elgar 2016) 350–352.
[138] Geiger and Izyumenko (n 9) 287.
[139] ibid 287.
[140] ibid 288.
[141] ibid 290.
[142] Senftleben, 'Flexibility Grave' (n 126) 763.
[143] Case C-469/17 *Funke Medien NRW GmbH v Bundesrepublik Deutschland* EU:C:2018:870, Opinion of AG Szpunar, para 40.

Directive, a so-called 'fair use' clause.[144] While not strictly corresponding to the US 'fair use' clause with its own four-factor system,[145] this flexible clause would function based on a balancing of fundamental rights.[146] This would not threaten the harmonisation of EU copyright law (and hence of phonogram protection), because this clause would be found in EU legislation itself. Meanwhile, it would allow more flexibility in the field of EU copyright law.[147] It would, moreover, permit an adaptation of the law when unforeseen societal, economic, or technological phenomena emerge.[148] Such an addition to the InfoSoc Directive would ameliorate the protection of artistic expression.

(iii) The Practical Inapplicability of the Quotation Exception for Most Artistic Uses

A third contribution by the CJEU concerns the quotation exception established in article 5(3)(d) of the InfoSoc Directive. According to that provision, reproduction and communication of a phonogram may be allowed by Member States when the phonogram is quoted in the context of a critique or review if: it relates to a phonogram that was already made lawfully available to the public; where the source, including the author's name, was made available if possible; where the work or related subject-matter was fairly used; and where it was used to the extent of the relevant purpose.[149] The Court found that this exception could only be relied on when the 'quoted' phonogram is identifiable in the quotation.[150]

Again, this conclusion was reached through the medium of a fair balance of fundamental rights. The Court first analyses the concept of quotation as to be understood in its usual meaning in everyday language. It thus applies the definition stated by the Advocate General in his Opinion, that is, the use by a user of an extract of a work or other subject-matter, or the full work or other subject-matter, to illustrate an assertion, defend an opinion, or compare the work or other subject-matter and the assertions of that user, and therefore to enter into a dialogue with the work or other subject matter.[151] Then, the Court reads article 5(3)(d) in the light of the freedom of the arts established in article 13 of the Charter.[152] It infers that the technique of sampling a phonogram can only fall under the quotation exception when there is an intention to enter into a dialogue,

[144] Geiger and Izyumenko (n 9) 301. See also Senftleben, 'Flexibility Grave' (n 126) 763. See also Martin Senftleben, 'Bridging the Differences between Copyright's Legal Traditions – The Emerging EC Fair Use Doctrine' (2010) 57 Journal of the Copyright Society of the USA 521, 541. On the US 'Fair Use' test, see Lyman Ray Patterson, 'Free Speech, Copyright, and Fair Use' (1987) 40 Vanderbilt Law Review 1. On the application of the US fair use test in cases of sound sampling, see Edward Lee, 'Fair Use Avoidance in Music Cases' (2018) 59 Boston College Law Review 1873.

[145] Geiger and Izyumenko (n 9) 302.

[146] ibid.

[147] ibid.

[148] Senftleben, 'Flexibility Grave' (n 126) 763.

[149] Copyright Directive, art 5(3)(d).

[150] *Pelham* (n 9) para 74.

[151] *Pelham*, Opinion of AG Szpunar (n 102) para 64.

[152] Charter, art 13; *Pelham* (n 9) para 72.

that the phonogram is recognisable to the ear, and that the already existing conditions under article 5(3)(d) are respected. Indeed, according to the Court, which follows the reasoning of the Advocate General, no dialogue is possible where the phonogram cannot be identified.[153]

Thus, the CJEU uses an internal limitation of EU copyright and related-rights law, the quotation exception, but reads it through the medium of fundamental rights to shape the internal limits of the system to ensure fair protection of artistic expression. As explicitly mentioned in *Spiegel Online*, the purpose of such use is to strike a fair balance between the right to freedom of expression of users of a work or other subject matter and the rights conferred on authors and other rightsholders and to ensure the actual use of the quotation exception when relevant.[154] Here, it can be said that this conclusion combines the flaws of the Court's criterion of what is 'recognisable to the ear' with those of an internal balancing, rather than an external one. Moreover, the obligation to have the intention to enter a dialogue, that the phonogram is recognisable to the ear, and that the already-existing conditions under article 5(3)(d) are fulfilled are inflexible and applicable only in a restricted number of cases, thus excluding typical sampling uses.[155] In the light of the foregoing, the Court's interpretation of article 5(3)(d) hardly protects artistic expression.

To sum up, phonogram protection is easy to acquire, thanks to easily satisfied conditions. It could be of interest to protect distinctive signs of goods or services offered by undertakings. While the CJEU tried to rule in favour of artistic expression by means of a balancing exercise of fundamental freedoms, the efficiency and flexibility of the solutions brought by the Court are restricted and could be favourable for such undertakings. Now that phonogram and sound mark protection have been individually analysed, it seems necessary to assess how the two systems may overlap in a way that endangers artistic expression (Section IV).

IV. The Overlap of Sound Mark Protection and Phonogram Protection as a Detriment to Artistic Expression

If sound mark protection is granted, phonogram protection will almost always overlap with it (Section IV.A), which will be detrimental to artistic expression (Section IV.B).

[153] *Pelham*, Opinion of AG Szpunar (n 102) para 64.
[154] *Spiegel Online* (n 113) para 82.
[155] Geiger and Izyumenko (n 9) 289–290.

A. The Overlap of Sound Mark Protection and Phonogram Protection

The protections of phonograms and sound marks clearly overlap for undertakings seeking protection for their sounds when used as signs to distinguish their goods or services. In particular, this would be relevant where the sound mark-protected content has been developed as part of brand lore.[156] The light and flexible criteria for phonogram protection would be easily, if not automatically, fulfilled when one has already obtained sound mark protection.

When it comes to substantive criteria, sound mark protection requires a distinctive sound,[157] this requirement being subject to the high threshold of resonance.[158] Meanwhile, phonogram protection is vested in a mere fixed sound that does not even need to be original.[159] A sound that met the substantive requirements for sound mark protection would thus always pass the threshold for phonogram protection. With regards to formalities, sound mark protection requires compliance with a burdensome registration process,[160] although the representation of the sound mark on the register has become more flexible and now allows the fixation of sounds as forms of representation.[161] Phonogram protection, on the other hand, generally requires no formality and, if national law provides some, these formalities are restricted to a mere inscription on commercialised copies of the phonogram.[162] Unless national law imposes this additional, albeit light formality, the formality requirements for phonogram protection would then be met when one has already recorded a sound and represented it as a sound file on a trade mark register. Finally, phonogram protection additionally requires a certain territorial connection,[163] but thanks to the concept of national treatment, this criterion is virtually fulfilled in most cases involving a Member State.[164]

It thus appears that, when sound mark protection is granted, phonogram protection will almost systematically overlap with it. The cumulation of the two protections can be detrimental to artistic expression (Section IV.B).

[156] Jane C Ginsburg and Irene Calboli, 'Intellectual Property in Transition: The Several Sides of Overlapping Copyright and Trademark Protection' in Niklas Bruun and others (eds), *Transition and Coherence in Intellectual Property Law: Essays in Honour of Annette Kur* (Cambridge University Press 2021) 312–315.

[157] Trade Mark Regulation, art 4(a); Trade Mark Directive, art 3(a).

[158] *Globo Comunicação* (n 22) para 45; *Ardagh Metal Beverage Holdings* (n 27) para 24.

[159] *Pelham*, Opinion of AG Szpunar (n 102) para 30.

[160] Trade Mark Regulation, art 6; Trade Mark Directive, art 1. On the criteria for registration, see *Sieckmann* (n 37) para 55. On the registration of non-traditional marks, see Gangjee (13) 66.

[161] Trade Mark Regulation, art 4; Trade Mark Directive, art 3.

[162] Rome Convention, art 11.

[163] Pila and Torremans (n 5) 270.

[164] Rome Convention, art 5.

B. The Detriment Posed to Artistic Expression with Expanded Sound Mark and Phonogram Protections

The threat to artistic expression posed by the overlap of phonogram protection and sound mark protection is twofold, because of the expansion of both sound mark and phonogram protection.

(i) The Detrimental Effect of the Expansion of Sound Mark Protection on Artistic Expression

The expansion of trade mark protection to sounds has adverse effects on artistic expression within the sound mark system itself. Sound mark proprietors enjoy a broad range of rights with claims for double identity, likelihood of confusion, and trade mark reputation to prevent artists from using identical, similar, or dissimilar sounds.[165] There are only a restricted number of limitations and defences on which artists could rely to avoid claims stemming from sound mark proprietors.[166]

More importantly, the expansion has eroded the internal limitations of the trade mark system aimed at excluding sounds not used in the course of trade to distinguish goods or services.[167] Combined with the lack of parallel development of users' rights regarding artistic uses,[168] non-commercial and mixed artistic users are now exposed to sound mark claims.[169]

The CJEU itself has acknowledged in early judgments that trade mark protection shall not be extended in a way that would clash with the public interest,[170] but has disregarded its own observations by eventually giving more weight to trade mark holders' interests.[171] Moreover, while the CJEU has engaged in discussing freedom of expression in EU trade mark law, this has been a marginal phenomenon.[172] Artistic expression plays a limited role in the trade mark system for the benefit of sound mark holders. Hence, within the sound mark system, the expansion of sound mark protection has adverse effects on artistic expression.

Outside the sound mark system, the expansion of trade mark protection within the natural realm of phonogram protection leads to the overlap of the two

[165] Trade Mark Regulation, art 9(2); Trade Mark Directive, art, 10(2).
[166] Trade Mark Regulation, arts 9(2) and 14(1); Trade Mark Directive, arts 10(2) and 14(1).
[167] Schovsbo (n 68) 564–566.
[168] ibid.
[169] ibid 557, 564–566.
[170] See for instance Joined Cases C-108/97 & C-109/97 *Windsurfing Chiemsee* EU:C:1999:230, [1999] ECR I-2779, para 25, as well as Case C-299/99 *Koninklijke Philips Electronics NV v Remington Consumer Products Ltd*. EU:C:2002:377, [2002] ECR I-5475, para 79. See Schovsbo (n 68) 565.
[171] Schovsbo (n 68) 566.
[172] Ilanah Simon Fhima, 'Trade Marks and Free Speech' (2013) 44 International Review of Intellectual Property and Competition Law 293, 319–320.

regimes, which translates into a clash of 'individual imperatives' between the purposes of phonogram protection and sound mark protection.[173]

Phonogram law, as a system of related rights, was developed jointly with copyright law in the EU.[174] As a result, its imperatives are derived from those of copyright law, namely the dissemination and preservation of cultural expressions,[175] including artistic expression. In that context, the main purpose of phonogram protection is to ensure the pursuit of the cultural innovation cycle, that is the 'incessant process of the creation of fresh, original human expression on the basis of pre-existing sources of inspiration'.[176] Cultural expressions are produced by the constant reinterpretation of pre-existing cultural artefacts.[177] Connected to this idea of innovation cycle is that of intergenerational equity. This idea can be explained in two steps. In the shorter term, it dictates that there should be synchronic intergenerational equity, that is intellectual property rules should permit co-existing generations of creators to share their experiences to together develop new cultural artefacts.[178] In the longer term, diachronic intergenerational equity ensures that future generations of creators shall be able to develop their own cultural artefacts based on previous generations' experiences.[179] Hence, to safeguard intergenerational equity, the phonogram regime should, in theory, require phonogram producers to allow artists to use their phonograms to create new ones.[180] This is possible *inter alia* by limiting the scope of IP rights.[181]

Meanwhile, the sound mark regime has diverging imperatives.[182] As part of the trade mark system, its core purpose is to safeguard 'market transparency'.[183] It permits the clear identification of the goods or services offered by an undertaking on the market so that consumers can individualise them and identify their commercial source.[184] This supports the prevalence of the public's favourite supplier of a type of good.[185] This is done by means of a stable distribution of intellectual resources among undertakings,[186] as evidenced by the possible

[173] Senftleben, *The Copyright / Trademark Interface* (n 10) 12. See Ginsburg and Calboli (n 156) 307–321.
[174] On the joint development of copyright and related rights, see Pila and Torremans (n 5) 222–233.
[175] Senftleben, *The Copyright / Trademark Interface* (n 10) 12.
[176] ibid 26, 29.
[177] Julie E Cohen, 'Copyright, Commodification, and Culture: Locating the Public Domain' in Lucie Guibault and P Bernt Hugenholtz (eds), *The Future of the Public Bomain: Identifying the Commons in Information Law* (Kluwer Law International 2006) 150. See also Jessica D. Litman, 'The Public Domain' (1990) 39 Emory Law Journal 965, 966–967.
[178] Shubha Ghosh, 'Why Intergenerational Equity' (2011) 2011 Wisconsin Law Review 103, 107.
[179] ibid 108.
[180] Senftleben, *The Copyright / Trademark Interface* (n 10) 35.
[181] Brett Frischmann and Mark P McKenna, 'Intergenerational Progress' (2011) 2011 Wisconsin Law Review 123, 125.
[182] Senftleben, *The Copyright / Trademark Interface* (n 10) 95.
[183] Annette Kur and Martin Senftleben, *European Trade Mark Law* (Oxford University Press, 2017) 6–8.
[184] Andrew Griffiths, 'A Law-and-Economics Perspective on Trade Marks' in Lionel Bently, Jennifer Davis and Jane C Ginsburg (eds), *Trade Marks and Brands: An Interdisciplinary Critique* (Cambridge University Press 2008) 245–255.
[185] Kur and Senftleben (n 183) 6.
[186] Senftleben, *The Copyright / Trademark Interface* (n 10) 95.

perpetual renewal of sound mark rights.[187] The CJEU has acknowledged the market transparency objective of trade mark protection.[188] It serves the defensive purpose of preventing competitors from using marks in a way that interferes with the communication of basic information about the commercial origin of goods and services.[189] This permits fair competition, consumer protection, and the proper functioning of markets.[190] Hence, the sound mark regime is inherently designed for commercial purposes, with little consideration for artistic matters.

Thus, while phonogram protection is guided by cultural imperatives demanding a perpetual cycle of creation with a constant flow of cultural resources, sound mark protection is characterised by trade imperatives demanding a static repartition of cultural resources.[191] These two types of imperatives diverge to such an extent that the expansion of sound mark protection within the artist-friendly realm of phonogram protection poses a bar to the cultural innovation cycle, putting artistic expression at threat of commercial claims.

(ii) The Detrimental Effect of the Expansion of Phonogram Producers' Rights on Artistic Expression

The expansion of phonogram producers' rights comes at the expense of artistic expression. The easily acquirable phonogram protection has experienced an expansion of its substantive scope of protection, which could be appropriated by sound mark holders trying to protect their sounds used as marks against artistic uses. The additional reliance on phonogram protection could permit undertakings to better protect their sounds used as marks, especially against other sounds that are not aimed at being signs used in the course of trade to distinguish their goods or services. This was rendered possible by the inadequate and inflexible solutions brought by the CJEU in *Pelham*.[192] The court failed to provide safeguards to artistic expression against phonogram producers' claims, which further strengthened the likelihood of successful reliance on the phonogram regime for undertakings.[193] This judgment has fundamental consequences on the innovation cycle and poses a bar to intergenerational equity. Future musicians are virtually prevented from using short pre-existing musical artefacts, thus putting artistic expression at threat.

While it is true that phonogram protection is only granted for a limited number of years,[194] as opposed to sound mark protection that can be renewed

[187] Trade Mark Regulation, art 52; Trade Mark Directive, art 48. See Pila and Torremans (n 5) 375.
[188] Case C-10/89 *SA CNL-SUCAL NV v HAG GF AG* EU:C:1990:359, [1990] ECR I-3711, para 13.
[189] Kur and Senftleben (n 183) 6.
[190] Senftleben, *The Copyright / Trademark Interface* (n 10) 134.
[191] ibid 95–96.
[192] *Pelham* (n 9) paras 39, 65, 74.
[193] Senftleben, 'Flexibility Grave' (n 126) 763.
[194] Directive 2006/116/EC of the European Parliament and of the Council of 12 December 2006 on the term of protection of copyright and certain related rights [2006] OJ L372/12, art 3(2).

indefinitely,[195] the cumulation of sound mark and phonogram protections for that period is still a powerful tool that could affect the very purpose of phonogram protection by posing a bar to the cultural innovation cycle. Moreover, during that timeframe, the phonogram could be updated to effectively start a new phonogram protection period.[196] Undertakings could combine their extensive rights under trade mark law with the extensive scope of protection available under the phonogram regime that goes beyond the 'course of trade' and 'in relation to goods or services' criteria, without any need for registration. This would only further put artistic expression at threat of commercial claims.

In sum, the cumulation of phonogram sound mark protections would be detrimental to artistic expression. This possibility would be against the imperatives of phonogram protection as it would affect the cultural innovation cycle.

(iii) Amending the Adverse Effects on Artistic Expression of the Overlap of the Phonogram and Sound Mark Regimes

To avoid the adverse effects of the overlap of sound mark and phonogram protections, it is necessary to amend the regimes. It has been proposed that we should rethink the grounds upon which to refuse registration, such as public order and morality, so that pre-existing signs with cultural significance would not be appropriated by undertakings.[197] Such a reform could effectively prevent the appropriation of major cultural artefacts by undertakings. However, it would not address the issue of undertakings gatekeeping their sounds used as marks to prevent artistic uses. These are usually not items of cultural significance that undertakings had appropriated but sounds that were composed with the original purpose of being marks used by undertakings.

To circumvent this problem, it has been argued that an *a posteriori* approach which balances the interests at stake should be adopted at the time of enforcement.[198] Rather than purely refusing to register non-traditional marks, like sounds, this approach would allow undertakings to protect their sounds used as signs to ensure market transparency while not affecting artistic expression. To reach that goal, it has been proposed to integrate a legal presumption of fair use in the honest practices test of trade mark exceptions to permit the use of a sound mark for artistic purposes.[199] Accordingly, the use in an artistic context of a sound mark would be presumed to be fair unless the rightsholder proves with

[195] Trade Mark Regulation, art 52; Trade Mark Directive, art 48.
[196] Ginsburg and Calboli (n 156) 314.
[197] See Martin Senftleben, 'Towards a New Copyright/Trademark Interface — Why (And How) Signs with Cultural Significance Should Be Kept Outside Trademark Law' in Graeme B Dinwoodie and Mark D Janis (eds), *Research Handbook on Trademark Law Reform* (Edward Elgar Publishing 2021).
[198] Gangjee (n 14) 80–88.
[199] Senftleben, 'Robustness Check' (n 64) 599–600.

individualised facts that unusual grievous harm has been inflicted on the sound mark.[200]

With regard to the rights of phonogram producers, it appears necessary to find a solution to avoid the outcome in *Pelham* in relation to artistic expression. The introduction of a fair use clause in the EU copyright regime could be a solution, as explained in detail above (see Section III.B.(ii)). This clause would protect the rights of phonogram producers while showing consideration to artistic expression by excluding fair artistic uses from the scope of their rights.[201]

To sum up, the cumulation of phonogram protection and sound mark protection would be detrimental to artistic expression. With the two regimes overlapping, undertakings could successfully rely on it to protect their sounds used as signs in the course of trade, at the expense of artistic expression. This possibility would be against the very imperatives of phonogram protection. Reforms to integrate fair use in both sound mark and phonogram regimes should be considered.

V. Conclusion

This article has tried to determine how sound mark and phonogram protections could overlap in EU law, and to examine the resulting cost upon artistic expression. First, it has analysed sound mark protection. While acquiring sound mark protection remains difficult, the criteria for registration have become more flexible and have been clarified. However, the rights granted to sound mark proprietors are extensive and allow them to challenge a variety of sounds, sometimes even outside the course of trade and with no relation to goods or services, at the expense of artistic expression. Second, it has demonstrated that phonogram protection is easy to acquire thanks to its light requirements. It has argued that the CJEU in *Pelham* failed to provide efficient and flexible safeguards to protect artistic expression against phonogram producers' claims. Finally, it has explored the detrimental effects on artistic expression caused by the overlap of sound mark protection and phonogram protection. It has shown that phonogram protection will easily, if not automatically, overlap with sound mark protection when the latter is granted. It has then demonstrated that the accumulation of the two protections leads to detrimental effects on the cultural innovation cycle, at the expense of artistic expression.

Today, European Union laws on phonogram protection and sound mark protection overlap easily, if not automatically, once sound mark protection is granted. This allows a cumulation of the two protections, although they are in theory guided by diverging imperatives. The combined effects of the expansion of sound mark protection and the non-protective approach to artistic expression

[200] ibid 592.
[201] Geiger and Izyumenko (n 9) 301.

regarding phonogram protection in *Pelham* cause serious interference with artistic expression. Neither of the two regimes adequately protects artistic expression, and the possibility of combining them puts artistic expression at serious risk. They inhibit the cultural innovation cycle by affecting intergenerational equity.

Reforms are necessary to mitigate the adverse effects of the overlap of the sound mark and phonogram regimes on artistic expression. In the sound mark regime, it is necessary to introduce a presumption of fair use so that the honest practices requirement does not inhibit the cultural innovation cycle. Similarly, a fair use clause should be integrated in the copyright regime so that a more flexible approach towards the artistic use of sounds subject to phonogram protection would be adopted. Otherwise, future artists in the EU may not be as lucky as David Gilmour was in front of the French courts.