

**Robots from Venus,
Robots from Mars:
Legal personhood for autonomous AI**

Imogen Lucy Grace Rivers

St John's College

University of Oxford

A dissertation submitted for the degree of

Doctor of Philosophy

January 2026

Acknowledgments

To Alex. Supervisor extraordinaire.

To Rosy, Matt, Anna, Dominic and Katie. For joy and love and grace.

To Jonathan Mark Luke Mallet. For lighting up my whole world forever.

AMDG!

CONTENTS

ABSTRACT	6
INTRODUCTION.....	7
I.1. Overview	7
I.2. A challenge-based analysis of legal personhood	9
I.2.1. The overview.....	9
I.2.2. A challenge for rights.....	11
I.2.3. A challenge for responsibility?.....	12
PART ONE	15
1. The Robotic Sexual Violence Problem.....	16
1.1. Introduction	16
1.2. What are Sexual Autonomous Robots?.....	18
1.2.1. Physically Embodying AI.....	19
1.2.2. Communication and Interaction	20
1.2.3. Autonomous Capabilities	21
1.3. The Robotic Sexual Violence Problem	24
2. Sex, Robots and Political Liberalism	28
2.1. Liberal principles and robotic sexual violence	29
2.1.1. The Harm Principle.....	29
2.1.2. Legal Rights and Legal Wrongs	31
2.1.3. Legal Protections	34
2.1.4. The Right to Privacy.....	36
2.2. Was that (morally) good for you too?.....	38
2.2.1. Religion and natural law	38
2.2.2. Feminist Voices	41
2.2.3. A Question of Virtue.....	44
2.3. The Liberal Principle of Legitimacy.....	48
2.3.1. The Question of Political Liberalism	48
2.3.2. Political Conceptions of Justice	51
3. Robots and Reciprocity	54
3.1. The Personal and the Political.....	56
3.1.1. Justice Behind Closed Doors: A Question of Reciprocity.....	56
3.1.2. Beyond the Porn Wars: Rape Myths and the Reporting Crisis	60
3.1.3. Autonomous Pornography	64
3.2. Virtues of the Polity.....	67

3.2.1. When is virtue political?	67
3.2.2. Testimonial Justice as a Political Virtue.....	69
3.2.3. Conventions Concerning Discrimination	70
3.3. Options for implementation	73
3.4. Concluding Remarks.....	75
4. Robots and Persons.....	77
4.1. The Grounds of Legal Personhood	77
4.1.1. Personhood Legalism and Personhood Moralism	78
4.1.2. Legislative Integrity.....	82
4.1.3. Personhood Functionalism.....	85
4.2. Oh Brave New World: Towards a Legal Framework for SARs.....	89
4.2.1. The Legal Personhood of SARs	90
4.2.2. Robot Rights.....	93
4.2.3. Objections and Replies.....	96
4.3. Concluding Remarks.....	102
5. Summing Up	104
5.1. Closing Considerations	104
5.2. Conclusion.....	107
PART TWO	108
6. The Responsibility Gaps Thesis	109
6.1. Introduction.....	109
6.2. What are LAWs?.....	114
6.2.1. Lethal independence.....	114
6.2.2. Adaptability.....	116
6.2.3. Discretion.....	118
6.3. The Responsibility Gaps Thesis	120
6.3.2. The value of responsibility per se?	124
6.3.3. The instrumental value of responsibility?	127
6.3.4. Alternative interpretations of the Responsibility Gaps Thesis.....	130
7. Answerability and blameworthiness	132
7.1. Moral responsibility as answerability	133
7.1.1. No gaps, but abundance	134
7.1.2. Epistemic and control conditions	137
7.2. Look: no gaps!	142
7.2.1. The epistemic condition.....	142

7.2.2. The control condition	147
7.2.3. The causal condition	150
7.3. Someone to blame	156
8. Criminal and civil liability gaps	160
8.1. Criminal responsibility for Massacre	161
8.1.1. Criminal responsibility	162
8.1.2. Punishment is no good (in Massacre)	165
8.1.3. Criminal responsibility revisited	169
8.2. Keeping things civil	171
8.2.1. Reparatory moral duties for Massacre	171
8.2.2. Private law remedies	184
9. Persons and products	197
9.1. Vicarious liability at common law	198
9.1.1. The enterprise liability rationale	199
9.1.2. Beyond employment: relationships reinterpreted	204
9.1.3. Close connections	209
9.2. Legal personhood and liability for LAWs	212
9.2.1. Legal personhood: a brief recap	212
9.2.2. Legal personhood for LAWs	214
9.2.3. Liability for LAWs	216
9.3. Product liability for LAWs	220
9.3.1. Product liability and enterprise liability	221
9.3.2. Four reforms... ..	224
10. Closing remarks on the Responsibility Gaps Thesis	231
CONCLUSION	234
References	236
Bibliography	236
Case law	310

ABSTRACT

The debate surrounding the ascription of legal personhood—viz., the capacity to hold legal rights and/or to bear legal responsibilities—to autonomous AI systems has so far focused on developing an appropriate legal framework to regulate wrongs caused *by* those systems. This has generated a growing literature on so-called ‘responsibility gaps’ where an autonomous AI system causes harm for which no agent appears to bear responsibility.

Part One of this dissertation shifts the focus: I argue that there are contexts in which we need to develop an appropriate legal framework to regulate wrongs inflicted *on* robots. I focus on sexual autonomous robots (SARs)—viz., humanoid AI-based autonomous robots which enact sexual relationships with humans. The imminent emergence of this kind of sexual technology introduces a troubling new way to enact sexual violence. I articulate and defend the first politically liberal argument for the legal proscription of what I call ‘robotic sexual violence’. I show that the attribution of legal personhood to SARs, including legal rights against robotic sexual violence, is part of an appropriate legal response.

Part Two returns to the question of responsibility which has so exercised academic and public debate on AI ethics. I focus on lethal autonomous weapons (LAWs)—viz., weapons systems which can select and engage targets without human intervention. I argue that the legal personification of LAWs would form part of an appropriate legal response to the risk of lethal accidents which the imminent proliferation of LAWs introduces. The ascription of a legal duty to LAWs not to direct attacks against civilians is justified by appeal to the role which LAWs play as quasi-employees of a military enterprise. Additionally, I argue for reforms to strict product liability mechanisms, reflecting the role which LAWs play in the space of legal obligations as both ‘products’ and ‘persons’.

INTRODUCTION

I.1. Overview

This dissertation aims to make a philosophical contribution to certain law and policy debates about the appropriate response to artificial intelligence (AI).¹ The dissertation focuses on two spheres of concern, which are addressed in Part One and Part Two respectively. The first sphere is about the regulation of AI in intimate relationships; the second sphere is about the regulation of AI in war. These apparently disparate spheres share three significant commonalities.

Firstly, there is a congruence between the core problems with which I engage in each sphere. The fundamental nature of the problem in each case concerns violence or, more specifically, the appropriate response to apparent violence involving autonomous AI. For the first time, we are developing systems to ‘stand in’ for human persons in relationships at the most intimate and the most impersonal levels. Violence in relationships of either kind is (usually) regulated with some strictness. When it is not so regulated, we often call for change because we feel we have principled reasons for believing that violence should be a matter of legal and political concern. The issue of violence becomes rather less clear cut, however, when either the victim (Part One) or the perpetrator (Part Two) is substituted for an AI system. Indeed, where one or more of the relata in a relation of violence is replaced by an AI system, we begin to wonder whether the relation is violent at all and what the appropriate response might be from a law and policy perspective. This underlying question concerning the appropriate response to apparent violence involving autonomous AI has not so far been identified in the AI ethics literature; its identification and analysis in two parts in this dissertation shines light on a fruitful new direction for that literature.

¹ ‘Artificial intelligence’ is notoriously difficult to define. For an overview of ‘human-centric’ and ‘rationalistic’ definitions: Turner (2019: 7-22).

Secondly, the technology studied in each sphere will involve *narrow* rather than *general* AI. With broad brush-strokes, AI is the increasingly ubiquitous phenomenon that some artificial (e.g., computational or robotic) systems display cognitive and/or autonomous capabilities that are in relevant respects analogous to the cognitive and/or autonomous capabilities of humans or other intelligent animals.² I will not be discussing the kind of physically embodied AI which populates science fiction like the *Terminator* films or *RoboCop*. These kinds of depictions are imaginative constructions of that elusive *general* AI which is comparable in its sophistication, flexibility and generality to human intelligence. This is standardly contrasted with *narrow* AI, which involves the skilful pursuit of pre-set goals, is manifest in present technology and is presumed to lack sentience, that is, the ability to feel pleasure and pain.³ In this dissertation, I will only be discussing narrow AI.

Thirdly, the solution in each case centres on the following insight: to introduce entities without legal rights or legal responsibilities—in other words, without legal personhood—into formerly human-human relations threatens to undermine the integrity of law and policy, at least in some cases; the remedy, or part of it, is to legally personify certain AI systems. Whilst it is uncontroversial to suggest that *general* AI might be a candidate for legal personhood, this insight has never been developed at any length in the literature on *narrow* AI, nor in the attendant law and policy debates; its articulation and defence in this dissertation is therefore a novel and timely contribution.

² For instance, Russell and Norvig (2009) define the subject of AI as ‘the study of agents that receive percepts from the environment and perform actions’ (c.f. Laukyte 2014, 2017).

³ Amoroso and Tamburrini (2021a); List (2021: 1217-8). C.f. Bostrom’s (2012: 73) ‘orthogonality thesis’ that ‘intelligence and final goals are orthogonal axes along which possible agents can freely vary’. Sparrow (2007) contests this thesis; for a reply: Burri (2017).

I.2. A challenge-based analysis of legal personhood

This section adumbrates the methodology and substance of the dissertation. I adopt what I call a ‘challenge-based’ approach to AI ethics in the law (**I.2.1**), one which articulates and solves new challenges in the interpretation and ascription of legal rights (**I.2.2**) and legal responsibilities (**I.2.3**) in an age of autonomous AI technologies, as opposed to one which identifies general normative principles for the regulation of AI.

I.2.1. The overview

This dissertation explores what justice requires in a world where human relationships—in love and in war—are increasingly mediated through AI. The dissertation aims to answer the following two questions:

- (1) What principles might justify the ascription of legal rights to autonomous robots?
- (2) When an AI causes harm, how should we allocate legal responsibility?

I adopt the structure of an integrated thesis, analysing (1) and (2) in two respective parts. But the underlying theme is preserved: the ascription of legal rights and/or legal responsibilities is definitive of legal personhood; the determination of the law’s future agents and patients will therefore depend upon the answers to (1) and (2).

Because this is a philosophy thesis which seeks to contribute to academic, legal and political discussions, I contextualise (1) and (2) using specific technological developments which are currently sparking debate. Part One examines (1) in the context of sexual autonomous robots (SARs)—viz., humanoid AI-based autonomous robots which enact sexual relationships with humans. Part Two centres on (2) in the context of war, studying the allocation of responsibility for wrongful harms caused by lethal autonomous weapons (LAWs)—that is, AI-based ‘weapons system[s] that, once activated, can select and engage targets without further

intervention by a human operator’.⁴ Throughout, I emphasise the importance of the background context in the normative evaluation of AI systems. More specifically, I seek to put into relief that which is often confined to the background context in debates concerning the particular technological developments with which this dissertation is concerned. So, Part One examines how the development, sale and use of SARs may intensify existing epistemic injustices which undermine mechanisms for the enforcement of the fundamental rights of women and girls; Part Two argues that the appropriate normative evaluation of LAWs depends upon recognising the technology as both a product and a ‘quasi-employee’ of enterprises which impose characteristic risks on third parties. One merit of the challenge-based approach which I adopt is that its attention to specific challenges, including their background context, assists not only in the resolution of philosophical debates in the AI ethics literature but also in the development of specific solutions which may assist in the development of law and policy.

One drawback of the challenge-based approach is, however, that it necessitates certain concessions to the presumptions which underpin real-world law and policy debates, in an effort to speak to those debates. In Part One, this is less of a drawback because—as I elaborate on in that part—there are principled reasons founded in the idea of respect, as well as practical ones, for the development of a Rawlsian politically liberal argument for the regulation of SARs. However, in Part Two, this dissertation makes a number of assumptions about the ethics of war. *Inter alia* I assume for the sake of argument that it is possible at least in principle to have a ‘just war’; that such a war would comply with something like the traditional *jus ad bello* and *jus in bellum* principles enshrined in international law, especially international humanitarian law; and that a war is not rendered necessarily unjust by the very use of LAWs.⁵ In truth, the problems with this kind of framework loom large, particularly in light of the vast scale of civilian

⁴ US Department of Defense (2012).

⁵ For critical engagement with this latter presumption: Purves *et al.* (2015).

casualties in the conflicts which have unfolded in Ukraine and the Middle East whilst I have been writing this dissertation. These conflicts have highlighted with particular vividness the inadequacy of current AI-based target selection and engagement systems at achieving anything close to compliance with just war principles,⁶ not to mention the brazen inadequacy of those principles at protecting human rights.⁷

I make two points in defence of my approach. Firstly, I make no claim that there actually has been or will be a just war, with or without LAWs. Secondly, in international legal and political debate concerning the regulation of such systems, ‘IHL [i.e., international humanitarian law] is central [...] as there is consensus that [L]AWS can be deployed insofar as they abide by the IHL principles of necessity, proportionality, and distinction.’⁸ Indeed, the UN Group of Governmental Experts on LAWs has endorsed preliminary criteria for the legal permissibility of LAWs which centre on compliance with international humanitarian law.⁹ It is in order to contribute meaningfully to the ongoing debate concerning the appropriate legal and political response to the imminent emergence of LAWs that I make the stipulated assumptions. With this overview in place, I will briefly sketch the arguments of Parts One and Two.

I.2.2. A challenge for rights

The academic and public debate surrounding the ascription of legal personhood—viz., the capacity to hold legal rights and/or to bear legal responsibilities—to autonomous AI systems has so far focused on developing an appropriate legal framework to regulate wrongs caused *by* those systems. This has generated a growing literature on so-called ‘responsibility gaps’ where an autonomous AI system causes harm with respect to which there appears to be no agent who

⁶ Davies *et al.* (2023); McKernan and Davies (2024).

⁷ Eggert (2022a, 2022b).

⁸ Taddeo and Blanchard (2021: 2).

⁹ UN GGE (2024, 2025).

can appropriately be held responsible for what happened. Part One of this dissertation shifts the focus: I argue that there are contexts in which we need to develop an appropriate legal framework to regulate wrongs inflicted *on* those robots.

I focus on SARs because the imminent emergence of this kind of sexual technology introduces a troubling new way to enact sexual violence. Robotic sexual violence cases share the following general structure. There is some human agent A, a SAR B and some action ϕ such that A ϕ -s B and, if B were human, A's ϕ -ing B would be a sexual offence. What is the appropriate legal response to robotic sexual violence?

There appear to be two classes of responses to this question. On one hand, there are those who argue for the legal permissibility of robotic sexual violence by appeal to politically liberal principles. On the other hand, there are those who argue for the legal proscription of robotic sexual violence by appeal to moral conceptions of the good. My contribution to this literature is to articulate and defend an argument for the legal proscription of robotic sexual violence by appeal to politically liberal principles which all free and equal citizens in a pluralistic society can reasonably accept. In particular, I argue that there is a Rawlsian 'public reasons' justification for the legal proscription of robotic sexual violence as part of the cultivation of those political virtues which are required by reciprocity in the just society. I show that the attribution of legal personhood to SARs, including legal rights against robotic sexual violence, would be part of an appropriate legal response. I conclude with some reflections on the wider ramifications of my arguments for the AI ethics literature.

I.2.3. A challenge for responsibility?

Part Two of the dissertation returns to the question of responsibility which has so exercised the academic and public debate on AI systems. In particular, Part Two centres on (2) in the context

of war, studying the allocation of responsibility for harms caused by LAWs. I make three principal contributions to the literature.

Firstly, I argue for a reframing of the debate. So far, most people seem to think that AI systems and LAWs in particular are at risk of generating some kind of ‘gap’ in moral responsibility such that the system might be able to cause harm for which no human agent can appropriately be held morally responsible. My first contribution is to show that the very idea of a gap in moral responsibility is conceptually flawed and that arguments for the existence of gaps in moral responsibility often depend upon a failure to distinguish between two distinct senses of ‘responsibility’, one which I call ‘answerability’ to denote the idea of calling an agent to answer for, explain or justify their conduct and one which I call ‘blameworthiness’ to capture the idea of holding an agent liable to blame, sanction or punishment.

My second contribution to the literature is to articulate and defend an alternative interpretation of the idea of a ‘gap’ in responsibility. I argue that there is some discrepancy between the (degree of) legal responsibility and the (degree of) moral responsibility which is attributable following lethal accidents involving LAWs. Contrary to popular misconception, the ‘gap’ appertains not to the criminal law remedies which are available in lethal accidents cases, but rather to the private law remedies. I contend that participants in enterprises for the production, sale and use of a LAW bear potentially significant reparatory moral duties when it goes wrong. However, existing private law mechanisms do not reflect those reparatory moral duties. So, I highlight a novel ‘responsibility gap’: when lethal accidents involving LAWs occur how, in principle, can we give legal expression and/or enforceability to those reparatory moral duties which accrue to participants in enterprises for their production, sale and use?

My third contribution is to argue for a novel solution to this ‘responsibility gap’ which I have identified under international private law. Firstly, I argue for the application of vicarious

liability in lethal accidents cases to ensure the legal enforcement and expression of those reparatory moral duties which attach to participants in the enterprise(s) which deployed the LAW. A corollary of this is that LAWs should be legally personified and endowed with a legal duty not to direct attacks against non-combatants. I prove that corollary. Finally, I argue for targeted amendments to strict product liability to ensure the legal enforcement and expression of those reparatory moral duties which attach to participants in enterprises for the production and sale of LAWs. I conclude with some reflections on this dissertation's overall contribution to philosophical, legal and political debates concerning the appropriate response to autonomous AI.

PART ONE

PUBLIC REASONS BEHIND CLOSED DOORS: LEGAL RIGHTS FOR SEXUAL AUTONOMOUS ROBOTS

*Do you believe in the human heart? I don't mean simply the organ, obviously. I'm speaking in the poetic sense. The human heart. Do you think there is such a thing?*¹⁰

¹⁰ Ishiguro (2021: 242).

1. The Robotic Sexual Violence Problem

*a dissociation of sensibility set in, from which we have never recovered*¹¹

1.1. Introduction

In 2017, the European Parliament called upon the EU Commission to table a legislative proposal laying down a set of civil law rules on robotics and AI. The most striking recommendation in the Parliament’s final report is that of

‘creating a specific legal status for robots [...] so that at least the most sophisticated autonomous robots could be established as having the status of electronic persons responsible for making good any damage they may cause [...]’.¹²

The Parliament’s proposal faced substantive criticism and was eventually abandoned.¹³ But legal personhood—that is, the (legal capacity for the) holding of legal rights and/or bearing of legal duties¹⁴—for autonomous robots has been a live topic of debate in the philosophical literature for several decades.¹⁵ In particular, in unison with the Parliament’s recommendation, a recent surge in academic interest in ascribing legal personhood to autonomous robots reflects and responds to the widespread concern that autonomous robots will generate morally and

¹¹ Eliot (1921).

¹² Delvaux (2017: 18).

¹³ For criticism of the Parliament’s proposal: Nevejans (2016); European Economic and Social Committee (2017); Karner (2019); Wagner (2019); Riehm (2020); Eidenmüller and Wagner (2021). Evidence of the proposal’s abandon appears in the European Parliament Resolution (2020/2014[2021]: ¶7).

¹⁴ There are different ways to cash out this orthodox view (see, generally, Smith 1928; Kurki 2019). These include the bearing of rights *or* duties (Gray 1997); the bearing of rights *and* duties (Bilchitz 2009: 42; *Nonhuman Rights Project v. Lavery* [2014] 124 A.D.3d 148, 998 N.Y.S.2d 248); the *capacity* for bearing rights/duties (Wise 2010: 17; *Restatement (Third) of Agency* 2006: §1.04; c.f. Rochfeld 2013: 15; Larenz 1983: 41) or for participating in legal relations (Savigny 1840, 1867, 1884; Wise 1998: 801; Salmond 1920). See also Kelsen’s minimalist approach to legal personhood: Kelsen (1919, 2005, 1945). I assume the orthodox view in this thesis, although it is not without its dissidents (see, e.g., Kurki 2019).

¹⁵ The *locus classicus* is Solum’s landmark (1992). But the earliest academic mention of this kind of approach appears in Stone (1972[2010]: n.26). Legal personhood for AI systems is also discussed in Koops *et al.* (2010), critiqued in Bryson *et al.* (2017) and Santoni de Sio and Mecacci (2021), but advocated for in Turner (2019) and Chesterman (2020). In the media, see, for example: Hern (2017).

legally problematic ‘responsibility gaps’, namely cases in which an autonomous robot causes harm or damage for which no human agent can justifiably be held responsible.¹⁶

In Part One of this dissertation, I shift the focus of this discussion in the following way. I develop a new argument for the ascription of legal personhood to certain kinds of autonomous robots. But my argument focuses not upon what these robots might do to *humans*, but rather upon what humans might do to *them*. To advocate for the ascription of legal *rights*, rather than legal *responsibilities*, to a certain class of autonomous robots is therefore my ultimate aim in Part One of this dissertation.¹⁷ I focus on SARs because the recent literature reveals that the imminent emergence of this kind of sexual technology introduces a troubling new way to enact sexual violence, one which calls for the development of a legal framework for their regulation.¹⁸ Part One of this dissertation is my contribution to that broader project.

In this chapter, I foreground the challenge posed by what I call the ‘robotic sexual violence problem’. In essence: what is the appropriate legal response to robotic sexual violence cases?¹⁹ In chapter two, I explore two classes of responses to the robotic sexual violence problem which have emerged in the SARs literature. On one hand, there are those who argue *against* the legal proscription of robotic sexual violence by appeal to politically liberal principles which all free and equal citizens in a pluralistic society can reasonably accept. On the other hand, there are those who argue *for* the legal proscription of robotic sexual violence

¹⁶ Matthias’ (2004) notion of ‘responsibility gaps’ has been especially influential in debates surrounding LAWs. See, for example: Sparrow (2007, 2016) Asaro (2012, 2016); Crootof (2015); Horowitz and Scharre (2015); Roff (2013); Galliot (2015). But ‘responsibility gaps’ are considered to be a problem for autonomous systems more generally (Mittelstadt *et al.* 2016; Santoni de Sio & Mecacci 2021; Kiener 2022a, 2022b).

¹⁷ Rights and responsibilities are, however, intimately intertwined in law, as I return to in chapter five.

¹⁸ See, for example, the first consultation report by the Foundation for Responsible Robotics, entitled ‘Our Sexual Future with Robots’, calling for the urgent development of a legal framework for SARs (Sharkey *et al.* 2017; E&T editorial staff 2017).

¹⁹ Notable recent contributions specifically concerning the robotic sexual violence problem include: Chatterjee (2020); Danaher (2017a, 2017b); Sparrow (2017); Strikwerda (2017). The robotic sexual violence problem has garnered political attention too: the USA House of Representatives and the UK’s Crown Prosecution Service recently banned the importation of child-like SARs and sex dolls. Note that the robotic sexual violence problem is distinct from the questions of whether *all* sex with robots should count as rape (Eskens 2017) and what, if anything, we should do about SARs designed to look like a particular human (Lancaster 2021).

by appeal to some moral conception of the good. This illuminates a lacuna in the recent literature: no one has argued *for* the legal proscription of robotic sexual violence by appeal to politically liberal principles. It is to the remedy of this omission that I turn in chapter three. Here, I articulate and defend the first politically liberal argument for the legal proscription of robotic sexual violence: I contend that robotic sexual violence is incompatible with a core value of political liberalism, one which Rawls²⁰ calls ‘reciprocity’. In chapter four, I suggest that the ascription of legal personhood to SARs, including legal rights to be secure from robotic sexual violence, would be a fruitful starting point for the implementation of the legal proscription of robotic sexual violence. I conclude in chapter five with some reflections on the ramifications of my arguments for wider debates in the AI ethics literature. To begin with, though, I wish to clarify the parameters within which my arguments unfold.

1.2. What are Sexual Autonomous Robots?

SARs are members of the wider class of social robots—namely, those ‘physically embodied, autonomous agent[s] that communicate[...] and interact[...] with humans on a social level’²¹ and are ‘specifically designed to elicit’²² anthropomorphic responses from users. There are three aspects to this definition of social robots which I wish to clarify in the context of SARs because the latter, as I will explain, physically embody AI (1.2.1), communicate and interact (1.2.2) and behave autonomously (1.2.3) in highly distinctive ways. A preliminary note on terminology: in this dissertation, I mention but will not use the terms ‘sex robot’ and ‘sexbot’ because these terms are used in academic and public discourse to refer ambiguously to sex dolls and SARs. But, as I explain in this chapter, the differences between sex dolls and SARs mean that SARs generate distinctive regulatory challenges.

²⁰ Rawls (1993[2005], 1997).

²¹ Darling (2014: 1-2).

²² *Ibid.*; c.f. Frank and Nyholm (2017: 307).

1.2.1. Physically Embodying AI

Firstly, SARs physically embody AI in a distinctive way. For one thing, SARs are humanoid and ‘designed to provide an artificial facsimile of a real human sex partner’.²³ As such, SARs are not sex toys, like vibrators or artificial vaginas. Nor will I explore the distinctive normative issues which animal, alien or generally non-humanoid robots might raise. The primitive prototypes for SARs currently available—known as sex dolls—are realistically shaped as human beings with silicone skin which is heated and textured to emulate human skin.²⁴ These are produced and sold by companies such as RealDoll—which sells customisable models of ‘male’ and ‘female’ sex dolls—and TrueCompanion—which has launched multiple editions of the ‘female’ sex doll ‘Roxxy’ and is allegedly developing a ‘male’ sex doll called ‘Rocky’.²⁵ Although the requisite technology, as I explain below, seems to exist and these prototypes are proceeding apace, I assume that SARs are not here yet, but soon will be—absent, of course, the kind of moratorium on their development, sale and use that campaigns like Richardson’s *Campaign Against Sex Robots* calls for.²⁶

It is important to emphasise that, whilst SARs do not yet exist, the anticipation of their arrival is by no means excessively futuristic. In particular, SARs embody the kind of AI technology which either presently exists or soon will, that is, narrow rather than general AI. As such, SARs will be designed to replicate human behaviour only in certain (e.g. sexual and social) interactions and, importantly, they will lack sentience—that is, the ability to feel pleasure and pain. Let us turn next to the question of what SARs will be able to do.

²³ Danaher (2017a: 73). There is some controversy as to whether sexual activity with SARs may properly be called sex (Odlind and Richardson 2023: 1-8). I will remain neutral as to such terminological disputes through abjuring use of terms like ‘sex’ when describing the use of SARs for sexual gratification.

²⁴ Sharkey *et al.* (2017); Sparrow (2017).

²⁵ Eskens (2017: 64).

²⁶ Richardson (2015, 2016, 2023). This campaign was recently rebranded as the Campaign Against Porn Robots for reasons which I discuss in chapter three.

1.2.2. Communication and Interaction

Recall that social robots ‘communicate[...] and interact[...] with humans on a social level’.²⁷

There are two salient aspects to emphasise in regard to this capability when it comes to SARs. The first has to do with that humanoid, physical embodiment which I have just discussed: SARs will communicate and interact in a humanlike way partly because they look human. The ‘body language’ of current sex dolls remains fairly primitive, including touch and movement detection, blinking, brow movements, head turning and head tilting.²⁸ But unlocking a more truly ‘social level’ of communication seems to depend upon the mobilisation of certain emotional responses in the user—remorse, grief or sympathy, say²⁹—which depend for their activation upon ‘bodies and faces with expressive capacities akin to those of the human form’.³⁰ By way of illustration, consider Sophia, a social humanoid robot developed by Hanson Robotics and (in)famous for being granted Saudi Arabian citizenship in 2017. Sophia has advanced capabilities in mimicking social communication and interaction, capabilities so advanced that ‘she’ allegedly induces strong feelings of ‘love and compassion’ in humans.³¹ As Di Sturco reports of studying Sophia:

‘She started to look at me and smile, and I looked at her, and at that point for me, she was not human, but there was kind of a connection’.³²

Physically embodied in ways which Sophia has proven possible, SARs would be able to engage a ‘social level’ of communication and interaction unprecedented in the sexual technology industry.

²⁷ Darling (2014: 1-2).

²⁸ Realbotix (2023); Galaitsi *et al.* (2019).

²⁹ Sparrow (2004: 102).

³⁰ *Ibid.*: 95; Gaita (1999: 269).

³¹ Greshko (2018); Hanson Robotics (2023).

³² Greshko (2018).

In addition to the consequences of their physical embodiment, conversation and interaction with SARs will be humanlike in its content and style in virtue of their AI equipment. The state-of-the-art in sex doll technology already employs AI to achieve some degree of humanlike conversational capability. For example, RealDoll's latest sex doll, 'Harmony', has conversational capabilities which encompass simple processes (e.g., 'remembering' facts) and more complex ones (e.g., engaging with 'her' user's emotions).³³ Synthea Amatus' sex doll, 'Samantha', speaks around 6,000 sentences and displays modes tailored to different conversational contexts—including 'romantic', 'fun', 'analysis', 'entertainment', 'sleep' and 'sexual' modes.³⁴ In 'her' 'family' mode, Samantha is 'programmed to talk about philosophy, animals, offer motivational advice and tell jokes'.³⁵ Meanwhile, it is alleged that TrueCompanion's Roxxxy

'knows your name, your likes and dislikes, can carry on a discussion and express her love to you and be your loving friend. She can talk to you, listen to you and feel your touch. She can even have an orgasm!'³⁶

But SARs will attain a more sophisticated level of communication and interaction. This is because SARs will 'learn', as I explain below.

1.2.3. Autonomous Capabilities

Unlike sex dolls, SARs have autonomous capabilities, that is, 'they can function without human causal intervention after they have been designed for a substantial portion of behaviours'.³⁷ For example, although Roxxxy is physically embodied and communicates on something

³³ Shen (2019). Controversially, Samantha has been designed with a lifelike Scottish accent because this allegedly sounds the least robotic (Murray 2017).

³⁴ Devlin (2020).

³⁵ Racco (2018).

³⁶ Hines (2009). There are truth-in-advertising failings to be noted here, especially concerning the capacity of a sex doll to 'know' or to be a 'friend', let alone a 'loving' one (Sharkey *et al.* 2017; Sparrow 2019).

³⁷ Anderson and Anderson (2011: 1).

approaching a social level, this sex doll is not a SAR because, in addition to other limitations, a user must position the doll's limbs themselves in order to interact with Roxxy. To qualify as autonomous, SARs will need to be able to physically conduct themselves in sexual situations with significantly less human causal intervention than Roxxy requires. Indeed, it is standard to distinguish three levels of autonomy in the general context of autonomous robotics: human-in-the-loop, human-on-the-loop and human-out-of-the-loop, where only human-on-the-loop or human-out-of-the-loop systems standardly count as *autonomous*. This is because, after activation, human-in-the-loop systems require human input in order to carry out the task for which the system is designed. This is why Roxxy is not properly called autonomous. Human-on-the-loop systems, by contrast, carry out their task without further human input after activation, although a human monitor can override or intervene. For example, so-called level-three autonomous vehicles—a level colloquially known as ‘eyes off’ the road—monitor the environment and control the vehicle but the human driver is expected to respond to a request to intervene.³⁸ Human-out-of-the-loop systems dispense even with the human monitor. Abyss Creations have built a sex doll with something like the required level of physical autonomy but the doll remains primitive in other respects.³⁹ Throughout this dissertation, I will understand autonomy as human-on-the-loop or human-out-of-the-loop.

Of course, you might object, it takes two to tango: in one sense, humans must remain ‘in the loop’ in any human-robot sexual interaction. But I take human-on-the-loop or human-out-of-the-loop autonomy to mean that a SAR can behave, in rough terms, like an—albeit uncanny—human sexual partner, able to initiate or respond to sexual advances. How would SARs develop this kind of advanced humanlike behaviour? In order to understand the level of autonomous communication and interaction which recent advances in AI have made possible,

³⁸ SAE (2021).

³⁹ Kleeman (2020); Sturges (2020).

it will help to invoke the distinction between so-called ‘top-down’ and ‘bottom-up’ AI. In top-down tools, we programme general principles about codes and norms of speech and conduct, as well as situation-specific information, into the AI; then, the algorithm infers a judgment about the appropriate response to that situation.⁴⁰ Sophisticated sex dolls currently employ top-down AI. Top-down AI predictably makes decisions in accordance with its input principles: we can even programme these systems to implement Kantian deontology⁴¹ and preference utilitarianism.⁴² But the problems with top-down tools are well-known: in particular, even if we can agree on which to implement, principles about codes and norms of speech and conduct may be vague, conflicting or highly context-specific.⁴³

In order to achieve the kind of situation-specific flexibility that human practical reasoning displays, then, I suggest that social robots and SARs will involve at least some bottom-up AI, in particular, a machine learning technique known as ‘deep learning’ which constructs artificial neural networks to mimic the structure and function of the human brain.⁴⁴ It is *deep* because it uses numerous hidden layers of nonlinear processing to extract features from data and to transform the data into different levels of abstraction.⁴⁵ Reinforcement *learning* occurs when the performance of the machine improves as it analyses more data.⁴⁶ Once a system has been trained on sufficient data, the idea is that it will output humanlike communication and interaction in social situations without ever having to be programmed with contentious general principles. It is likely that advanced large language models such as

⁴⁰ Sinnott-Armstrong and Skorburg (2021: 4). A primitive example: Asimov’s (1950) three laws of robotics.

⁴¹ Arkoudas *et al.* (2005).

⁴² Oesterheld (2016). Programming consequentialist principles (e.g. minimise suffering) runs into problems: minimising suffering, say, could require the annihilation of humans (Bostrom 2014); information about future consequences is also difficult to estimate. Gips (1995) attempts to overcome these problems, but without much success (Wallach and Allen 2009: 86-90).

⁴³ Sinnott-Armstrong and Skorburg (2021: 5-6). Anderson and Anderson (2011) attempt to resolve these problems, but others remain pessimistic (Arkin 2009: 106-108; Wallach and Allen 2009: 127-129).

⁴⁴ The *Commission Guidelines on the definition of an artificial intelligence system established by Regulation (EU) 2024/1689 (AI Act)* note that deep learning ‘is a technology behind many recent breakthroughs in AI’ (European Commission 2025: (38)).

⁴⁵ Sinnott-Armstrong and Skorburg (2021: 4).

⁴⁶ Deep Learning Glossary, Deep AI.

ChatGPT employ a blend of top-down and bottom-up AI. Similarly, then, SARs would be trained to respond in fluid and humanlike ways to sexual situations using vast data sets of human-human sexual scenarios. One well-known problem with this kind of bottom-up AI, which I return to in chapter three, is the difficulty of ensuring that the data sets do not ingrain problematic biases into the SAR's responses.⁴⁷ Another problem, as I explain in the next section, is that there is an inescapable unpredictability which the flexibility of deep learning introduces.⁴⁸ To some extent, my discussion in Part One of this dissertation is an exploration of one way in which these two problems intersect. Let us turn, then, from these primarily descriptive matters to the normative ones which will preoccupy us for the rest of Part One of this dissertation.

1.3. The Robotic Sexual Violence Problem

As Sharkey and others highlight, the imminent emergence of SARs raises a number of pressing normative issues concerning their regulation in law.⁴⁹ Part One of this dissertation focuses on the robotic sexual violence problem, which has preoccupied the nascent literature on SARs.⁵⁰ As we can see from the foregoing advertisement for Roxxy, SARs appeal to consumers in part because they promise to alleviate loneliness. For instance, Levy famously argued that widely available SARs could curb the loneliness epidemic in industrialised countries and even curtail harmful practices in the sex industry.⁵¹ In addition, Jecker sets forth a 'dignity-based argument for affording older people access to sex robots as part of reasonable efforts to support their

⁴⁷ For an illuminating discussion of bias in bottom-up AI systems: Cohen (m.s.).

⁴⁸ The learning process is often *unsupervised* as the algorithm discovers patterns in a data set without the programmer imposing any theory or categories in advance (Wood, Deep AI). But the process may also be supervised. In particular, Anthropic (2023) suggests that a set of principles—based on documents as diverse as the Universal Declaration of Human Rights and Apple's Terms of Service—may be used to train deep learning systems in ways which would *mitigate* if not *remove* the harms associated with the unpredictability of their learning processes.

⁴⁹ Sharkey *et al.* (2017). For a helpful overview of the moral and legal issues raised by SARs generally, see texts in Danaher and McArthur (2017). Feminist critiques of SARs are collected in Richardson and Odland (2023).

⁵⁰ See *supra* n.19.

⁵¹ Levy (2007).

central human capabilities’.⁵² The idea is that SARs would help people who are otherwise unable to do so to access the health and emotional goods of sexual activity.⁵³ Indeed, Bianchi has argued that ‘sex robots may serve as a beneficial tool to enable sexual functioning for older adults with disabilities [and, in particular,] to overcome some of the hurdles that exist in relation to sex, dementia, and consent’.⁵⁴ But, as I explain in this section, SARs also generate a troubling new way to enact sexual violence, which calls us to reflect on the kinds of legal principles which should govern the use of SARs.

It is important to clarify at the outset that I am not talking about the enactment of sexual violence scenarios with an entity which has been purposefully programmed to withhold consent for the purposes of the user’s sexual gratification. The sex doll Roxxy, by TrueCompanion, has several preprogrammed personality types—‘Frigid Farah’, ‘Young Yoko’, ‘Wild Wendy’, ‘S&M Susan’ or ‘Mature Martha’.⁵⁵ ‘Frigid Farah’, described by the manufacturer as a ‘very reserved’ personality that ‘does not always like to engage in intimate activities’,⁵⁶ has provoked criticism from commentators who believe that this mode facilitates the simulation of coercive or non-consensual sexual conduct.⁵⁷ In the context of sex dolls, of course, there is an easy fix, namely to programme sex dolls to always consent enthusiastically to their users’ sexual advances.⁵⁸ But the enactment of sexual violence with SARs cannot be so easily resolved. Let’s see why this is.

⁵² Jecker (2020: 26).

⁵³ C.f. McArthur (2017).

⁵⁴ Bianchi (2020: 38).

⁵⁵ Kleeman (2017).

⁵⁶ Brown (2019: 112).

⁵⁷ Danaher (2017a); Timmins (2017).

⁵⁸ The attempt to produce SARs displaying this kind of sexual enthusiasm raises other concerns. As Sparrow (2017) highlights, this might seem to propagate ‘the idea that women are always available for sex [which] is itself a central tenet of rape culture’ (475, 468). On this, see: Buchwald *et al.* (2005); Herman (1988). In a similar vein, Gutiu writes that very sexually available SARs would ‘embed the idea that women are passive, ever-consenting sex objects, and teach users that when getting consent from a woman, ‘only no means no’ (2012: 15). On this and other harmful stereotypes which SARs would seem to contribute to, see also: Gutiu (2016); Sparrow (2017); MacWilliam (2023).

One crucial difference between SARs and sophisticated sex dolls, as I explained in **1.2.3**, is that, to adequately mimic human behaviour in a vast array of different situations, SARs are expected to involve some form of deep learning. One upshot of this is that SARs will behave in ways which are *in principle* impossible for the manufacturer, seller or user of a SAR to predict. Matthias was the first to draw our attention to this intrinsic unpredictability which accrues to *learning* in AI systems:

‘[R]ecently there are machines in development or already in use which are able to decide on a course of action and to act without human intervention. The rules by which they act are not fixed during the production process, but can be changed during the operation of the machine, *by the machine itself*. This is what we call machine learning.’⁵⁹

As a result, even a tried-and-tested SAR, which reliably seems to welcome its user’s sexual advances in all test cases, may behave unexpectedly in any new situation. Call it ‘consenting’ when a SAR gives the verbal or behavioural signs which standardly indicate consent to sexual activity when they are given by a human (e.g., the SAR verbalises an enthusiastic ‘Yes!’). Call it ‘withholding consent’ when a SAR does not consent.⁶⁰ Because they are autonomous, in other words, SARs will sometimes withhold consent in response to users’ sexual advances.⁶¹

Of course, this is analogous to sexual activity between human beings: a person who has always responded enthusiastically to their partner’s sexual advances in the past may withhold

⁵⁹ Matthias (2004: 177).

⁶⁰ You might worry that it does violence to our sexual language to apply the concepts of ‘consensual’ or ‘non-consensual’ to sexual activity with SARs since these concepts appropriately apply only to activities between humans. I defer to the wider literature on robotic sexual violence here, where it is standard in setting up the problem to consider robotic sexual violence cases as normatively significant simulacra of sexual violence. See, e.g., Danaher (2017a); Sparrow (2017).

⁶¹ Of course, this is not the only troubling way in which SARs might behave unpredictably. For instance, a SAR could also be unexpectedly sexually coercive towards the user, a scenario which raises questions concerning the appropriate allocation of responsibility for wrongful harms caused by a SAR. Because I am primarily concerned, as I explained in **1.1**, with the ascription of *rights* rather than *responsibilities* to autonomous robots, I defer further reflection on this possibility until chapter five.

consent to sexual activity on any new occasion. This may be unpredictable and for no externally discernible reason. The distinctive problem with SARs is that, when *they* withhold consent, there seems to be a morally problematic ‘rights gap’ because, whilst robotic sexual violence cases are intuitively troubling, no one’s legal rights are violated if the user ignores the SAR’s signals and proceeds with sexual activity regardless. One prominent concern, therefore, is that SARs would seem to create troubling ways of enacting sexual violence which are not (yet) proscribed by law. Consider, by way of illustration, the following case:

Robotic Rape: A SAR, B, withholds consent in situation, S. In S, an adult human, A, intentionally penetrates the ‘vagina’, ‘anus’ or ‘mouth’ of B with A’s penis. A does not reasonably believe that B consents in S.

Robotic Rape is a robotic sexual violence case: if B were a person, A’s penetrating B would constitute a criminal offence under English law, namely rape.⁶² The robotic sexual violence problem consists in the following question: what is an appropriate legal response to robotic sexual violence cases like Robotic Rape?

⁶² Sexual Offences Act 2003 c.42(1).

2. Sex, Robots and Political Liberalism

*Let us live, my Lesbia, let us love,
and all the words of the old, and so moral,
may they be worth less than nothing to us!*⁶³

There appear to be two classes of responses to the robotic sexual violence problem in the SARs literature. On the one hand, there are those who believe that the state may not proscribe conduct simply on the basis of some moral conception of the good; rather, laws should be justified by appeal to principles acceptable to all citizens. In this camp, we find those of a politically liberal inclination who believe that robotic sexual violence is among those harmless private sexual activities with which the law may not legitimately interfere (2.1). On the other hand, there are those who believe that the state *may* appropriately proscribe conduct which is not compatible with a certain moral conception of the good.⁶⁴ Here, we uncover an unlikely alliance between religious, feminist and virtue theoretic perspectives, which converge on the view that the appropriate response to the robotic sexual violence problem is to legally proscribe robotic sexual violence or, indeed, the production, sale and use of SARs altogether (2.2). In 2.3, I propose a novel solution to this dispute: why not, I suggest, seek to justify the legal proscription of robotic sexual violence by appeal to ‘public reasons’ which are acceptable to all reasonable citizens?⁶⁵ It is to the development of such an argument that I turn in the next chapter.

⁶³ Catullus 5.

⁶⁴ This is a theme widely discussed in the literature on neutralism and perfectionism in politics. Neutralists endorse some version of the principle of state neutrality—broadly, the claim that state action should be neutral among rival moral conceptions of the good (Dworkin 1978; Ackerman 1980; Larmore 1987; Rawls [1993]2005). Perfectionists reject that principle. I believe that this terminology of neutralism and perfectionism will not be illuminating in the present context: for one thing, there are lots of different ways to cash out the principle of state neutrality (Wall 2021); for another thing, as I explain in the following chapters, whilst Rawls’ political liberalism is standardly viewed as neutralist, it seems to be compatible with *some* kinds of state action to promote a conception of the good.

⁶⁵ The idea of public reason has roots in the work of Hobbes, Kant and Rousseau but has become increasingly influential in contemporary normative philosophy as a result of its development in the work of Rawls ([1993]2005,

2.1. Liberal principles and robotic sexual violence

In this section, I consider three standard rationales for legally proscribing conduct in liberal societies, one which appeals to considerations of harm (2.1.1), one which invokes basic principles of justice (2.1.2) and one which explores whether SARs might be eligible for certain kinds of legal protections (2.1.3). None of these rationales seems to justify the legal proscription of robotic sexual violence. Because of this, McArthur suggests that robotic sexual violence is one of those private sexual activities with which the law should not seek to interfere (2.1.4).⁶⁶ Note: there has been some speculation as to whether engagement in robotic sexual violence might serve a therapeutic or cathartic purpose for those with certain kinds of deviant or dangerous sexual desires.⁶⁷ However, as Sparrow explains, there is no empirical basis for these claims and the requisite data acquisition appears effectively infeasible.⁶⁸

2.1.1. The Harm Principle

One rationale for legally proscribing robotic sexual violence might appeal to some variant of Mill's harm principle, which holds that '[t]he only purpose for which power can be rightfully exercised over any member of a civilised community, against [her] will, is to prevent harm to others'.⁶⁹ In what sense, then, might robotic sexual violence harm—or risk harming—others?

With broad brush-strokes, harm may be caused 'directly' or 'indirectly' to others. For example, if a child were made to observe Robotic Rape this would directly cause harm; indeed, it would seem to constitute child sexual abuse under English law.⁷⁰ But to make matters stark,

1997, 1999a, 1999b, 2001), Habermas (1990, 1995, 1996, 1998) and Gaus (1996, 2009, 2011, 2012, 2016, 2017). In this thesis, I restrict my scope to the Rawlsian way of cashing out this idea.

⁶⁶ McArthur (2017).

⁶⁷ For example, in July 2014, the roboticist Ronald Arkin suggested that child-like SARs might be used to 'treat' paedophilic tendencies in the same way that methadone is used to treat heroin addiction.

⁶⁸ Sparrow (2019).

⁶⁹ Mill (2011[2019]: 17). There are well-known problems concerning the definition of harm. For various interpretations of Mill's concept of 'harm': Gray (1996: 57); Brink (2008: 42); Donner (2008: 161); Dyzenhaus (1992: 546). I gloss over these debates here.

⁷⁰ NSPCC (2023).

I will bracket these kinds of direct harms to others by assuming that robotic sexual violence is a private sexual activity involving only consenting adults. So: would robotic sexual violence cause, or risk causing, indirect harm to others? The obvious worry is that enactment of robotic sexual violence in private would make people more likely to enact sexual violence on human beings too. Gutiu has advanced an argument like this for the legal proscription of robotic sexual violence.⁷¹ Since SARs do not yet exist, we might seek empirical support for this claim in the media effects and pornography literatures. But the claim that exposure to or enjoyment of representations of violent activities is correlated with a likelihood to enact their real-world equivalents is heavily contested.⁷² Similarly, there is a long history of controversy about the relationship between the use of pornography depicting sexual violence and the perpetration of real-world sexual violence,⁷³ with no emerging consensus.⁷⁴ So, I propose that we set aside these kinds of harm-based arguments for the legal proscription of robotic sexual violence. This is not to say, though, that the pornography literature has nothing to teach us concerning the appropriate regulation of robotic sexual violence. Quite the contrary, as I explain in the next chapter when I defend an alternative, more robust argument for the legal proscription of robotic sexual violence. My argument will turn on the idea that exposure to or enjoyment of representations of violent activities can produce and reproduce a particular kind of *epistemic injustice* without necessarily *harming* anyone.

⁷¹ Gutiu (2012).

⁷² See, for example, Commonwealth of Australia (2009); Jenkins (2012).

⁷³ For excellent overviews of the last forty years of debate: Mikkola (2019: ch.2); Srinivasan (2021: ch.2); Wright, Tokunaga and Kraus (2016: 183).

⁷⁴ In a recent meta-analysis on the connection between pornography and the perpetration of sexual violence, Ferguson and Hartley (2020) found no correlation in the case of nonviolent pornography and, although there was slight positive correlation for violent pornography, ‘the current evidence was unable to distinguish between a selection effect as compared to a socialization effect’. The empirics of this debate are deeply vexed and I take no stand on them here.

2.1.2. Legal Rights and Legal Wrongs

In their (2000) discussion, Gardner and Shute argue that in sexual violence cases, it is in principle possible to violate important moral and legal rights without actually harming anyone. As such, one might infer, the legal proscription of some private sexual activity should depend primarily not upon whether it is harmful, but upon whether or not it violates moral or legal rights. Although I will challenge it in chapter four, there is a general presumption that SARs lack moral and legal personhood.⁷⁵ Because of this, to justify the legal proscription of robotic sexual violence, we would need to show that it violates the moral or legal rights of some human person.⁷⁶ To see why this will not seem to work, consider the composition and foundations of legal rights.

Standard accounts of legal rights have two central components, one which analyses the *form* of legal rights and one which analyses their *function*. Hohfeld's analysis of legal rights is by far the most widely followed and influential systematisation of the form of rights.⁷⁷ Writing specifically on the law, Hohfeld distils four different legal positions residing under the term 'right' as it is used in law: *privilege*, *claim*, *power*, and *immunity*. Claims and privileges are first-order norms, that is, they determine, respectively, what the relevant parties *ought* to do or *may* do; powers and immunities are second-order norms—that is, they are norms governing other norms—which determine, respectively, whether parties are *able* or *unable* to change other norms.⁷⁸ Specifying the function of a legal right—in other words, determining the

⁷⁵ See, e.g., Danaher (2017a: 88), who 'assum[es] throughout [...] that [SARs] are not persons', and Sparrow (2017: 467): '[t]here is [...] one scenario in which it *would* make sense to be concerned with the consent of a robot to sexual relations, which I mention only to set aside: when (hypothetical) artificially intelligent robots have become 'moral persons''.

⁷⁶ Of course, if SARs were, say, rented out by a corporation, we would also need to consider whether robotic sexual violence violates the moral or legal rights of an 'artificial person', namely that corporation which owns the SAR. I set aside this possibility here.

⁷⁷ Hohfeld (1919: 35–64).

⁷⁸ Sumner (1987: 27–29). For elaborations of the Hohfeldian analysis: Fitzgerald (1966: 221–233); Wellman (1985: 7–15); Sumner (1987: 18–31); Jones (1994: 12–25); Kramer (1998: 7–60); Edmundson (2004: 87–102); Wenar (2005: 224–237, 2005[2021]); and Harel (2005: 192–193).

necessary and sufficient conditions for some (composition of) Hohfeldian incident(s) to constitute a legal right—has been a major source of controversy within analytic legal philosophy for many years. Most theories can be classified as will theories or interest theories.⁷⁹ Will theories have been developed and refined by a number of authors;⁸⁰ interest theories by, most notably, Raz,⁸¹ MacCormick⁸² and Kramer.⁸³ I will briefly gloss the characteristic features of these theories:

- **Will theories**

Definition: for some candidate right-holder A, some (composition of) Hohfeldian incident(s) is a legal right of A's if A holds legal powers to annul, waive, enforce, or transfer the correlative duties.⁸⁴

Example: within limits and subject to various exceptions, holders of private law rights (e.g. legal rights in the law of contract, property, torts, or trusts) are empowered to waive compliance with certain obligations owed to them, forgo remedial rights, and annul or transfer some such claims, as well as certain private law powers, liberties, and immunities.⁸⁵

Value: will theories emphasise the affinity between rights and values that are related to an individual's control over his/her life, including personal liberty, self-realisation, agency, and autonomy.⁸⁶

⁷⁹ Some, like Vallentyne (2007), have sought to advance alternative or 'hybrid' theories of the function of legal rights which go beyond the interest/will theory stalemate. There is good reason, however, to think that these theories are merely versions of the interest theory (Frydrych 2018). As such, I set aside these theories for reasons of scope.

⁸⁰ Hart (1982); Steiner (1994); Simmonds (1998); Wellman (1995).

⁸¹ Raz (1986, 1994).

⁸² MacCormick (1977).

⁸³ Kramer (2001, 2013, 2017).

⁸⁴ Hart (1982: 183–4, 1983: 35); Steiner (1994). Steiner (1994) points out that only claims and immunities have correlative duties. I set this restriction aside because, since *compositions* of Hohfeldian incidents may also be rights, there is room for privileges and powers within the will-theoretic framework too.

⁸⁵ See, e.g., Stevens (2007: 17–19).

⁸⁶ Kramer (1998: 75); Sumner (1987: 47).

- **Interest theories**

Definition: for some candidate right-holder A, some (composition of) Hohfeldian incident(s) is a legal right of A's if an aspect of A's well-being is a sufficient reason for holding some other person(s) to be under a correlative duty.⁸⁷

Example: your claim that I should not wilfully strike you is a legal right because some aspect of your well-being (e.g. your health, bodily integrity) is a sufficient reason for holding me to be under a duty not to wilfully strike you.

Value: interest theories tap into the intuitively plausible connection between holding rights and well-being.⁸⁸

But according to either of these two main accounts of legal rights, there does not appear to be anyone whose autonomy or interests tell in favour of the legal proscription of robotic sexual violence. In the case of will theories, this is because, on its face, there is no legal person who appropriately should hold legal powers to annul, waive, enforce, or transfer a duty on A to refrain from ϕ -ing B. Quite the contrary: the only legal person whose autonomy is at stake in robotic sexual violence cases like Robotic Rape would appear to be A and the best way to protect A's autonomy would seem to be to legally enshrine their immunity from interference in their private sexual activities. In the case of interest theories, on its face, there is no legal person such that an aspect of their well-being is a sufficient reason for holding A to be under a legal duty not to ϕ B. I defer further examination of the rights question until the chapter four, but this discussion suffices for now to suggest that robotic sexual violence violates no person's legal rights.

⁸⁷ Raz (1986: 166, 1994: 266–269). Unlike will theorists, interest theorists tend to emphasis claim-rights more than powers (Kramer 2013: 246, 2017: 49). Still, holding liberties, powers, and immunities is often also in the service of one's interests (MacCormick 1977: 205).

⁸⁸ Interest theories seem to be more capacious than will theories, able to accommodate both unwaivable rights (since the possession of these is typically good for right-holders, as examples in criminal and constitutional law show) and the rights of children and people with certain disabilities (who have interests that rights can protect but arguably lack the kind of autonomy which will theory rights seem, at least at first blush, to depend upon). This is why I focus on interest theories in chapter four.

2.1.3. Legal Protections

Of course, you might point out, we sometimes legally proscribe conduct which is neither harmful to other persons nor in violation of any person's legal rights. For example, in most jurisdictions, animals and natural objects are not recognized as legal persons and, as a result, cannot bear legal rights.⁸⁹ Still, certain kinds of cruel or destructive treatment of them are often legally proscribed. Perhaps analogous kinds of legal protections, short of legal rights, might apply to SARs themselves.

One author who has explored this kind of line of argument is Darling.⁹⁰ She explores the emerging literature on the normative valence of different human relationships with social robots.⁹¹ Primitive social robots include interactive robotic toys like Sony's AIBO dog, Innvo Labs' robotic dinosaur Pleo and the therapeutic Paro baby seal.⁹² Psychological research documents that we form emotional attachments to these robots that are surprisingly strong and often affect how we believe they ought to be treated.⁹³ Pace Sparrow⁹⁴—who assumes that 'we would be forced to reassess our attitude' upon realisation that a social robot was not sentient—our projection of emotional states and even moral entitlements onto social robots does not seem to depend upon belief in their actual sentience. Owners of Sony AIBO dogs, for example, while fully aware of its non-sentience, instinctively ascribe mental states to their artificial companion

⁸⁹ Well-known exceptions include New Zealand's Whanganui River and the Ganges and Yamuna rivers in India, which were all granted legal personhood in 2017. The Ecuadorian Constitution (arts. 10, 71–74) recognises the rights of ecosystems to exist and flourish, gives people the authority to petition on behalf of ecosystems, and requires the government to remedy violations of these rights. In *Nonhuman Rights Project v. Lavery* [2014], the New York State Supreme Court ruled that animals were not legal persons and, as such, could not bear legal rights. On this, see Kurki (2019: ch.1).

⁹⁰ Darling (2014).

⁹¹ Hegel *et al.* (2009). On our tendency to anthropomorphise social robots and what this means from a regulatory perspective: Bartneck *et al.* (2007); Carpenter (2013); Duffy (2003); Kahn *et al.* (2012); Knight (2014); Saebeck and Bartneck (2010).

⁹² For an insightful discussion of social robots, including Reben's BlabDroids: Campbell (2020: ch.2).

⁹³ Turkle (2010); Darling (2014). When the US military began testing a robot that defused landmines by stepping on them, the commander called off the exercise on the basis that it was 'inhumane' (Garreau 2007; c.f. Carpenter 2013 and Kolb 2012).

⁹⁴ Sparrow (2004: 103-104).

and feel guilty for ‘mistreating’ it.⁹⁵ As Darling writes, ‘violent behavior towards robotic objects *feels* wrong to many of us, even if we know that the ‘abused’ object does not experience anything’.⁹⁶ She also notes that ‘successful societal pushes for animal abuse laws have followed popular sentiment rather than consistent biological criteria’.⁹⁷ In a similar way, then, she predicts that legal protections for social robots will emerge on the basis of the affective contents of our relationships with them, rather than on the basis of any particular principled criteria.

At first blush, my arguments in Part One of this dissertation might seem to complement Darling’s claims, offering a principled justification for that ascription of legal protections to SARs which Darling predicts but does not justify. However, there are three reasons why I set aside Darling’s legal protections framework. Firstly, there are important differences between animals and SARs, which suggest that *principled* rationales for legally protecting animals may not apply to SARs. One such rationale grounds legal protections against animal cruelty in the fact that animals experience pleasure and pain. This is not the case for SARs which, as I explained in chapter one, lack sentience. According to this way of thinking, laws proscribing animal cruelty seem justified on grounds that do not apply in robotic sexual violence cases. Secondly, there are also *pragmatic* rationales for legally protecting certain natural objects which do not extend to SARs. For instance, prohibitions on environmental degradation are typically justified instrumentally by appeal to human interests in, say, the health benefits of a clean environment.⁹⁸ But I have already set aside the possibility that robotic sexual violence causes harms like this to human interests: robotic sexual violence is not, say, unhealthy for

⁹⁵ Friedman *et al.* (2003: 276-277).

⁹⁶ Darling (2014: 13). As Levinas (1989) and Lingis (1994) explore, *looking human* seems to be an important basis for the development of normative relationships. It is worth noting, though, that Mori (1970) hypothesised that a person’s affective response to a humanlike robot would abruptly shift from empathy to revulsion as the robot approached, but failed to attain, a truly human appearance. This descent into eeriness is known as ‘uncanny valley’.

⁹⁷ Darling (2014: 17).

⁹⁸ Given the difficulties of enforcing environmental protections without appealing to the legal rights of any legal persons, Stone (1972[2010]) has famously argued for the ascription of legal personhood to trees and other natural objects.

others in the way that pollution is, so why prohibit the former? Finally, Darling’s arguments for legal protections are predicated on the claim that “rights for robots” [...] assumes a futuristic world of fully autonomous and highly sophisticated androids that are nearly indistinguishable from humans’.⁹⁹ In chapter four, I show that her discarding of the legal rights framework for autonomous robots here is undermotivated. The considerations advanced so far in this section are by no means decisive, but I believe that they give sufficient reason to take my next argument seriously.

2.1.4. The Right to Privacy

McArthur advances an argument which seems to suggest that robotic sexual violence is among those private sexual activities which the law should not interfere with.¹⁰⁰ Amongst the activities standardly protected by the right to privacy in liberal democracies are sexual activities where the only persons involved are consenting adults. Both the UN Human Rights Committee and the European Court of Human Rights have treated sexual life as an integral part of one’s privacy and, in a series of landmark judgments, ruled that laws prohibiting same-sex sexual activities constitute an unjustifiable interference with the right to respect for private life. In *Dudgeon v. UK* [1981]¹⁰¹ the European Court found criminal laws prohibiting consensual same-sex sexual activities between adults violated the right to respect for private life. The court went on in *Norris v. Ireland* [1988]¹⁰² to state that ‘[a]lthough members of the public who regard homosexuality as immoral may be shocked, offended or disturbed by the commission by others of private homosexual acts, this cannot on its own warrant the application of penal sanctions when it is consenting adults alone who are involved’.¹⁰³ Could the same kind of reasoning not

⁹⁹ Darling (2014: 1, n.4).

¹⁰⁰ McArthur (2017: §3.1).

¹⁰¹ [1981] ECHR 22 Oct 1981.

¹⁰² [1988] 6/1987/129/180.

¹⁰³ C.f. *Modinos v. Cyprus* [1993] 16 EHRR 485; *Toonen v. Australia* [1994] Communication No.488/1992, 4/4/1994).

apply to robotic sexual violence too? Shocking, offensive or disturbing private sexual activities, ‘when it is consenting adults alone who are involved’, do not usually warrant legal proscription.

Indeed, on the assumption that SARs are not legal persons, the use of SARs generally might even appear to be legally analogous to the use of sex toys, which is protected by the right to privacy in liberal democracies. This connection is most explicit in the ruling by a federal court in *Reliable Cons. v. Earle*,¹⁰⁴ which overturned Texas’ law against ‘obscene devices’ (sex toys). This judgment cited the US Supreme Court’s ruling regarding same-sex sexual activities in *Lawrence v. Texas* that ‘adults may choose to enter upon [a sexual] relationship in the confines of their homes and their own private lives’.¹⁰⁵ By the same token, in regard to sex toys, in *Reliable Cons. v. Earle* the court held that ‘[t]he right the Court recognized [in *Lawrence*] was not simply a right to engage in the sexual act itself, but instead a right to be free from governmental intrusion regarding ‘the most private human contact, sexual behaviour’’.¹⁰⁶ According to this line of reasoning, arguments against robotic sexual violence might seem misguided from the outset, ‘since they presume the legitimacy of something that is in fact illegitimate, [namely] the interference by society in the lives of individuals, in matters of strictly private concern’.¹⁰⁷

Of course, there are limits to this principle. In the UK, for example, the Domestic Abuse Act 2021 clarifies the law by restating, in statute, the broad legal principles established in the case of *R v Brown*:¹⁰⁸ a person commits an offence if they cause serious harm or, by extension, death to another person and the occurrence of this in the context of consensual ‘rough sex’ is

¹⁰⁴ 517 F.3d 738 (*No. 1*) (5th Cir. 2008); 538 F.3d 355 (*No. 2*) (5th Cir. 2008).

¹⁰⁵ *Lawrence v. Texas* (2003) 539 U.S. 558. C.f. *Griswold v. Connecticut* [1965] 381 U.S. 479; *Twyman v. Twyman* [1993] 855 S.W.2d 619).

¹⁰⁶ *Reliable Cons. v. Earle* (*No. 1*) at 744.

¹⁰⁷ McArthur (2017: 46). For critical elaboration of this line of argument: McArthur (2017: §3.1).

¹⁰⁸ *R v Brown* [1993] 2 All ER 75 House of Lords.

not a defence. But, as I have explained, *a fortiori* robotic sexual violence is a harmless sexual activity involving only consenting adults: what basis could there be for its legal proscription?

2.2. Was that (morally) good for you too?

In this section, I gloss three kinds of arguments which may be offered in favour of legally proscribing robotic sexual violence or, indeed, the production, sale and use of SARs altogether. In such a short space, the arguments against these competitors cannot, of course, be decisive. I wish only to motivate my subsequent appeal to the idea of public reason by evincing the sense that there is something disrespectful about the attempt to justify legal proscriptions on private sexual conduct by appeal to the kinds of reasons which not everyone can reasonably accept.

2.2.1. Religion and natural law

Herzfeld explores the use of SARs from religious perspectives, concluding that, at least from a Judeo-Christian point of view, the use of SARs would be ‘a form of idolatry’.¹⁰⁹ This ‘idolatry’ of ‘replacing relationship with one another with relationship to a machine’ is connected to a morally problematic dissociation of love and sexual desire:

‘Love ‘made to measure’ is not love, for then we would have tamed it, taken away its wildness and mystery, as well as its demand that we give ourselves to each other, perhaps even unto death.’¹¹⁰

Could this kind of Judeo-Christian evaluation of the use of SARs inform their legal regulation?

¹⁰⁹ Herzfeld (2017: 134). C.f., Fernández *et al.* (2025: [56]-[63]).

¹¹⁰ *Ibid.*

If you are a certain kind of natural law theorist, in particular one sympathetic to Finnis,¹¹¹ you might say yes.¹¹² A Finnisian natural law argument for the legal proscription of robotic sexual violence might go as follows:¹¹³

- (i) There is a *pro tanto* reason to legally proscribe any conduct if (a) that conduct is not in accordance with practical reason as given by natural law and (b) the legal proscription of that conduct would help people to live in accordance with practical reason;¹¹⁴
- (ii) (a) The enactment of robotic sexual violence is not in accordance with practical reason as given by natural law and (b) the legal proscription of that conduct would help people to live in accordance with practical reason;

So,

- (iii) There is a *pro tanto* reason to legally proscribe the enactment of robotic sexual violence.
- (i) is a standard tenet of natural law theories, on which positive law derives its moral authority from its ability to guide subjects to live in accord with an underlying ‘natural law’. As Finnis writes:

‘[p]ositive law [...], for its relevance in practical reason and the judgments of conscience, [...] depends upon natural law with its master principle of ‘love of neighbor as oneself’ and its more specific principles identifying kinds of act compatible with

¹¹¹ Finnis (1981, 2008, 2011a).

¹¹² It is beyond my scope to engage with another aim of Finnis (2008), namely the rejection of same-sex marriage from the point of view of natural law theory. I am, however, sympathetic to the view articulated by Macedo (1995: 280) that ‘[t]he focus on procreation [in Finnis’ account of same-sex marriage] appears opportunistic: selected so as to allow sterile heterosexuals into the tent while keeping homosexuals out’. For further discussion: Finnis (1997b: 37-43).

¹¹³ Note that the natural law argument is stronger than it needs to be since it would motivate a legal proscription on the use of SARs altogether, and not just the legal proscription of robotic sexual violence.

¹¹⁴ The reason is *pro tanto* because, as Finnis (1981: 351) explains, it is only the ‘central’ case of law and the ‘focal’ sense of legal validity which necessitate a strong connection between positive law and natural law in order for positive law to retain legal validity or moral authority. In borderline cases of positive law and secondary senses of legal validity, conduct which is not in accordance with practical reason may be validly legally permissible. As to how positive law retains moral authority in these cases: see Tan (2002).

intelligent love of human persons and their fulfilment, and kinds incompatible with it.¹¹⁵

So: would robotic sexual violence be compatible with intelligent love of human persons and their fulfilment?

Clearly not according to Herzfeld's interpretation of Judeo-Christian sexual ethics. Another way to defend (ii)(a) appeals to the concept of marriage as 'an intrinsic good'.¹¹⁶ Finnis defines 'morally bad kinds of sex act' as ones which 'set[...] the wills of the choosers [...] against the good of marriage'.¹¹⁷ This would encompass robotic sexual violence (and, generally, the use of SARs) because such acts are intrinsically devoid of 'openness to procreation'.¹¹⁸ But marriage is so integral to human fulfilment for Finnis that 'we are directed to that good [of marriage] by a first principle of practical reason'.¹¹⁹ So robotic sexual violence (and, generally, the use of SARs) would be contrary to the requirements of practical reason as given by natural law. Thus, (ii)(a). Would the legal proscription of robotic sexual violence help? Since marriage is usually a requirement of practical reason, 'the intelligibility and worth of its contours are bases for rejecting some legal arrangements and definitions and mandating others'.¹²⁰ Natural law theorists might then defend the legal proscription of robotic sexual violence as an efficacious way to combat 'the development of practices and ideologies that truly tend to undermine the sound understanding and practice of marriage'.¹²¹ Hence, (ii)(b). (iii) derives validly from the former premises.

There are three principal threads of objection to this kind of argument for the legal proscription of robotic sexual violence. Firstly, one might reject (ii)(a), denying, for example,

¹¹⁵ Finnis (2008: 401).

¹¹⁶ *Ibid.*: 389.

¹¹⁷ *Ibid.*: 393.

¹¹⁸ *Ibid.*

¹¹⁹ *Ibid.*

¹²⁰ *Ibid.*: 390.

¹²¹ George (2004: 85).

the Judeo-Christian perspective or the account of natural law which judges that robotic sexual violence would be morally wrong because it dissociates sex from love and/or procreation. But, secondly, even granting this starting point, we might deny (ii)(b): as Rawls explains, there is a problem of ‘futility’ which attends the attempt to enforce a certain moral conception of the good—such as Judeo-Christian sexual ethics—under conditions of ‘reasonable pluralism’, that is, the fact that ‘[t]he political culture of a democratic society is always marked by a diversity of opposing and irreconcilable religious, philosophical, and moral doctrines’, of which some are ‘perfectly reasonable’.¹²² Of course, what Rawls calls ‘practical’ stability might still be attainable through political persuasion and comprehensive enforcement. Thirdly, however, ‘[f]inding a stable conception [of justice] is not simply a matter of avoiding futility’;¹²³ rather, in a democratic society, we should ‘have stability for the right reasons’.¹²⁴ For Rawls, this occurs within an ‘overlapping consensus’, where citizens agree on principles of justice wholeheartedly from within their own perspectives. Why is this stability for the right reasons? I return to this question below, but one way to think about it is in terms of a certain kind of morally valuable respect. As Solum writes, it appears that ‘respect for citizens as free and equal does mean that we should give our fellow citizens the sort of reasons that they could reasonably accept’.¹²⁵ It seems disrespectful, then, to justify the legal proscription of robotic sexual violence by appeal to its incompatibility with Judeo-Christian sexual ethics or an associated account of natural law.

2.2.2. Feminist Voices

One concern which you might have with respect to the foregoing considerations is that the effort to justify law by appeal to principles of justice which all citizens can reasonably accept

¹²² Rawls ([1993]2005: 141-142, 3-4).

¹²³ *Ibid*: 142.

¹²⁴ Rawls (1997: 781).

¹²⁵ Solum (1992: 736).

may seem at odds with the advancement of justice for historically marginalised communities.¹²⁶ Indeed, many feminists believe that political liberalism is therefore intrinsically limited in its ability to advance gender justice: Baehr ‘concede[s] that a [politically liberal] feminism [...] will not be a radical doctrine’;¹²⁷ Brennan grants that ‘it will lack the kind of revolutionary potential associated with feminist political theories’¹²⁸ and Lloyd likens advancing principles of gender justice within a Rawlsian framework to ‘operating with one hand tied behind one’s back’.¹²⁹ Moreover, Yuracko contends that feminism is intrinsically at odds with the idea of public reason because feminist political claims are grounded in particular conceptions of human flourishing, that is, in particular accounts ‘of what a good life looks like’.¹³⁰ If it is true that gender justice is incompatible with the idea of public reason, what would it mean for the law governing SARs?

Feminists have long argued that the cause of gender justice calls for radical upheaval in the law governing sexual ethics.¹³¹ In a similar vein, in 2015, Richardson founded the Campaign Against Sex Robots, which argues for a moratorium on the development, sale and use of sex dolls and SARs ‘for the dignity of women and girls’.¹³² Pace Levy,¹³³ Sandberg¹³⁴ and Devlin,¹³⁵ Richardson argues that these devices represent not a progressive development in human relationships but rather the continued dehumanisation of women and girls. In particular, she believes that the production, sale and use of sex dolls and SARs would ‘contribute to gendered inequalities found in the sex industry’ and ‘reinforce relations of power

¹²⁶ Mills (2017: chs.8-9).

¹²⁷ Baehr (2008: 208).

¹²⁸ Brennan (2004: 94).

¹²⁹ Lloyd (1998: 210).

¹³⁰ Yuracko (2003: 9).

¹³¹ MacKinnon (1989, 2006, 2016); Langton (1993, 2009).

¹³² Richardson and Odland (2023: ch.1).

¹³³ Levy (2007).

¹³⁴ Sandberg (2015).

¹³⁵ Devlin (2015).

that do not recognise [women] as human subjects'.¹³⁶ Richardson's account is explicitly grounded in a particular conception of human flourishing which she calls a 'politics of love',¹³⁷ and describes as 'central to [her] work and the work of the campaign'.¹³⁸ The politics of love 'demand[s] that mutuality be the hallmark of sexuality' rather than what Richardson calls 'property relations'.¹³⁹ But SARs—as manufactured, sold and used objects—entrench 'property relations' rather than 'mutuality' in the context of human sexuality: '[p]eople are not property, nor is property capable of anything other than fictional personification'.¹⁴⁰ Richardson believes that the politics of love is central to the achievement of gender justice and, since SARs are incompatible with this conception of human flourishing, their production, sale and use should be legally proscribed.

There are many criticisms in the SARs literature concerning the substance of Richardson's arguments. Danaher, for example, challenges the assumption that 'property relations' are intrinsically at odds with 'mutuality' in sex, drawing on the example of respectfully paying for sex within marriage as a cultural practice of the Merina tribe.¹⁴¹ But it is besides my purposes in Part One of this dissertation to evaluate the substance of Richardson's claims. Instead, I want to focus on the bigger picture because there are both practical and principled reasons why we as feminists should not be so quick to dismiss the idea of justifying laws by appeal to public reasons which everyone can reasonably accept. Practically, principles of justice which are justified by appeal to public reasons seem more likely to drive social change. As Baehr writes:

¹³⁶ Richardson (2015: 292, 2016, 2023); c.f. Morrigan (2023).

¹³⁷ Richardson (2023: 188).

¹³⁸ *Ibid.*: 185.

¹³⁹ *Ibid.*

¹⁴⁰ *Ibid.*

¹⁴¹ Danaher (2017b). The example of the Merina tribe is taken from Zelizer (1996) and discussed in Brennan and Jaworski (2015). On variations in social meanings regarding paying for certain goods: Sandel (2012).

‘[C]ompare saying the citizens’ access to violent pornography should be restricted because such pornography undermines women’s status as equal citizens, to saying that access to such pornography should be restricted because the human body should not be objectified.’¹⁴²

In arguing for the same political principle by appeal to ideas of equal citizenship which are ‘implicit in our political culture’, the former argument ‘make[s] feminist social change seem possible’.¹⁴³ In a similar way, in Part One of this dissertation I aim to defend the legal proscription of robotic sexual violence by appeal to principles which, as I show in the next chapter, are implicit in our public political culture and, as such, may drive effective feminist social change. But there are principled reasons, too, why I appeal to the idea of public reason in this thesis. In recent years, feminist scholars have changed their tone somewhat, arguing that robust principles of gender justice *are* compatible with political liberalism.¹⁴⁴ But in the next chapter, I go further, arguing that robust principles of gender justice are in fact *intrinsic* to the Rawlsian idea of public reason.¹⁴⁵ These principles—entailments of the political relation between free and equal citizens—motivate the legal proscription of robotic sexual violence.

2.2.3. A Question of Virtue

There is increasing interest in the AI ethics literature concerning how virtue ethics might inform the design of AI systems.¹⁴⁶ Vallor argues that the development of social robots should be informed by a virtue theoretic perspective:

¹⁴² Baehr (2008: 197).

¹⁴³ *Ibid.*

¹⁴⁴ See, for example, Baehr (2008); Chambers (2018: 48–55); Hartley and Watson (2010); Lloyd (1998); Neufeld and Van Schoelandt (2014); Schouten (2013, 2017, 2019); Watson and Hartley (2018).

¹⁴⁵ Watson and Hartley (2018) defend a new interpretation of reciprocity, which lies at the heart of the idea of public reason. Their interpretation involves robust principles of gender justice. But I argue that such principles may be derived from within a purely Rawlsian understanding of reciprocity.

¹⁴⁶ Tonkens (2012); Wallach and Allen (2009).

‘From the perspective of virtue ethics, people who spend [...] their free time getting immense pleasure from torturing robots [...] are not living well or flourishing [and] because a good human being embodies character traits and patterns of activity directly contrary to sadism, our robot abusers are less noble for their sadistic habits even if no living persons are injured by those habits.’¹⁴⁷

Peeters and Haselager apply this kind of thinking to the case of SARs, arguing that ‘a virtue ethical analysis of intimate human–robot relationships could inspire the design of robots that support the cultivation of virtues’.¹⁴⁸ If Vallor’s thinking is correct, robotic sexual violence would be condemned from the perspective of virtue ethics. But can the viciousness of some activity motivate its legal proscription?

According to Danaher,¹⁴⁹ the viciousness of robotic sexual violence offers some reason in favour of its legal proscription. The idea is that the person who enjoys robotic sexual violence enactment—somewhat like the person who laughs at racist jokes or enjoys racist artworks—expresses something about their moral character that is worthy of social condemnation. Specifically, he argues that ‘those who engage in acts of purely robotic rape and child sexual abuse either demonstrate an immoral desire for the real world equivalents of those acts, and/or a disturbing moral insensitivity to the social meaning of those acts’.¹⁵⁰ Danaher’s arguments for this claim derive from work done on the ethics of virtual acts of violence, that is, acts of violence performed within the world of a video game. First, Danaher draws on Luck’s work on the so-called ‘gamer’s dilemma’—broadly, the question of why, if neither act tells us anything about moral character, do many people believe that virtual acts of murder are less morally condemnable than virtual acts of child sexual abuse—to develop his argument that there is a

¹⁴⁷ Vallor (2016: 211).

¹⁴⁸ Peeters and Haselager (2021: 55). For example, Peeters and Haselager suggest that SARs equipped with ‘consent-modules’ could support the cultivation of compassion in supervised, therapeutic scenarios.

¹⁴⁹ Danaher (2017a).

¹⁵⁰ *Ibid*: 89.

connection between moral character and the enjoyment of certain kinds of representation.¹⁵¹ But Patridge has developed the case that virtual acts of sexual violence, including child sexual abuse, give us much better evidence of a deficient moral character.¹⁵² This is because virtual acts of sexual violence have an ‘incorrigible social meaning’—that is, there is a limit on the range of reasonable interpretations of these representations—and anyone who derives pleasure from these activities either displays a problematically underdeveloped moral sensitivity (by interpreting outside that range) *or* expresses a morally problematic character (which interprets within that range and still derives enjoyment). Drawing a natural analogy with the enactment of sexual violence using SARs, Danaher argues that there is an incorrigible social meaning which attaches to the enactment of robotic sexual violence, one which expresses something seriously deficient about a person’s moral character.

Whether the law may concern itself with the development of good moral character in its citizens has been the source of much dispute.¹⁵³ Danaher argues for the legal proscription of robotic sexual violence by appeal to so-called ‘legal moralist’ arguments that the justification for legally proscribing a given type of conduct is intrinsically tied to the moral wrongfulness of that type of conduct.¹⁵⁴ In particular, Wall contends that ‘[i]t is a proper function of the criminal law to promote good character and to restrain or discourage people from engaging in activities that cause moral harm to themselves or to others’.¹⁵⁵ Wall distinguishes four primary domains of moral concern: (i) autonomy (ii) well-being; (iii) moral character; and (iv) impersonal excellence. Traditional liberal theories of criminal law hold that the criminal law may only appropriately govern immoral conduct affecting (i) and, to some extent (ii), but

¹⁵¹ See: Luck (2009); Luck and Ellerby (2013).

¹⁵² Patridge (2010, 2013).

¹⁵³ C.f. the Hart/Devlin debate concerning whether the state should be permitted to use the criminal law to enforce ‘social morality’—viz., certain arrangements of customary norms that a society does in fact endorse and practice (Hart [1964]1968: 17-24). See: Hart ([1964]1968); Devlin ([1965]1968). A modern take on the debate is Green (2013).

¹⁵⁴ For an overview of different varieties of legal moralism: Duff (2014).

¹⁵⁵ Wall (2013a: 459).

certainly not (iii) since the latter would amount to ‘soulcraft’—viz., the use of law to foster good character—which is usually presumed to be incompatible with liberal theories of criminalisation.¹⁵⁶ Wall, however, argues that that same concern to respect subjects of the law as free and equal persons which animates liberal thought should also sustain a criminal law engagement with (iii). *In nuce*, this is because that concern for the individual subject which underlies concern for (i) and (ii) should also sustain a concern for her success in her rational aims, that is, in those aims that she has and does not have good reason to revise or abandon.¹⁵⁷ Wall goes on to argue that success in her rational aims depends upon a subject’s being committed to pursuing a sound conception of the good, a commitment which itself turns upon the goodness of her moral character.¹⁵⁸ If Wall is right, Danaher argues, this would seem to support the legal proscription of robotic sexual violence.

The principal problem with Danaher’s argument for the legal proscription of robotic sexual violence is that it depends upon a particular conception of good moral character.¹⁵⁹ As we have seen, there are practical and principled problems which attend the attempt to legally proscribe individual conduct by appeal to principles such as these which not everyone can reasonably accept. Indeed, for Klosko governments refraining from coercing citizens on grounds that they can reasonably reject is ‘part of what it means to treat people with adequate respect’;¹⁶⁰ Quong concurs: ‘respect for persons requires that they not be politically coerced on grounds that they cannot reasonably accept’.¹⁶¹ In the next chapter, I will defend an alternative

¹⁵⁶ For a classic defence of this kind of position: Feinberg (1984). C.f. Husak (2007: 73-76). Sometimes concern for (i) clashes with concern for (ii), particularly if individual acts of autonomy impact on individual well-being (e.g. acts of self-harm). In fact, as Danaher (2017a: 78) points out, ‘this is the quintessential class animating debates between liberals/libertarians and paternalists’.

¹⁵⁷ Wall (2013a: 461-462); Scanlon (1998: 119, 126-143).

¹⁵⁸ Wall (2013a: 462-465).

¹⁵⁹ Another problem is that the argument rests on the idea that certain kinds of conduct have an incorrigible social meaning. But in subsequent work, Danaher (2017b) challenges this kind of ‘symbolic consequences argument’ on the basis of the contingency of social meanings.

¹⁶⁰ Klosko (2003: 169).

¹⁶¹ Quong (2004: 246). This emphasis on coercion is not without its critics in the political liberalism literature (Bird 2014; Quong 2013). I return to the question of coercive enforcement in the next chapter.

politically liberal argument for the legal proscription of robotic sexual violence. My argument will turn on the idea that the state may appropriately be concerned with the cultivation of certain kinds of virtue in its citizens but the virtues at issue must be *political*—that is, justified by appeal to public reasons—rather than merely *moral*—that is, justified by appeal to a moral conception of the good. Such considerations are not decisive, but I believe that they will suffice to explain why I wish to carve out an alternative justification for the legal proscription of robotic sexual violence.

2.3. The Liberal Principle of Legitimacy

The overarching problem with the arguments advanced in the previous section seems to be that free and equal people within a given political community, especially in modern liberal democracies, disagree about moral matters, particularly when it comes to sex. Given this deep and persistent pluralism, there is something disrespectful about justifying the legal proscription of robotic sexual violence by appeal to one conception of sexual morality which not everyone can accept. This prompts the following refinement of the robotic sexual violence problem: given a plurality of moral conceptions of the good among persons viewed as free and equal, which principles may appropriately guide the development of a particular legal framework for the regulation of SARs and, especially, robotic sexual violence?

2.3.1. The Question of Political Liberalism

The refinement of the robotic sexual violence problem echoes what I call the ‘question of political liberalism’ which Rawls sets out to answer in *Political Liberalism*:

‘[H]ow is it possible for there to exist over time a just and stable society of free and equal citizens, who remain profoundly divided by reasonable religious, philosophical, and moral doctrines?’¹⁶²

Exploring Rawls’ answer to this question will help us to answer the refinement of the robotic sexual violence problem since his starting points are similar to mine, seeking ‘fair terms of social cooperation between citizens regarded as free and equal’¹⁶³ subject to the fact of reasonable pluralism. Part of Rawls’ answer to the question of political liberalism is his ‘liberal principle of legitimacy’:¹⁶⁴

‘Our exercise of political power is fully proper only when it is exercised in accordance with a constitution the essentials of which all citizens as free and equal may reasonably be expected to endorse in the light of principles and ideals acceptable to their common human reason’.¹⁶⁵

He adds that

‘Only a political conception of justice that all citizens might be reasonably expected to endorse can serve as a basis of public reason and justification’.¹⁶⁶

In this subsection, I unpack the liberal principle of legitimacy in order to evince, in the next subsection, the criteria which a public reasons argument for the legal proscription of robotic sexual violence must satisfy.

Rawls’ liberal principle of legitimacy is premised on a conception of the constituency of public reason as comprising *reasonable persons*, viz., persons who accept two main ideas.

¹⁶² Rawls ([1993]2005: 4).

¹⁶³ *Ibid.*: 3.

¹⁶⁴ Rawls ([1993]2005: 225, 1995: 175ff.) distinguishes between legitimacy and justice: although they draw on the same set of political values, they have different domains and legitimacy makes weaker demands than justice. Exercises of political power may be legitimate but unjust, but the converse is not possible. I focus on legitimacy in this thesis because it gives a standard for the acceptability of exercises of political power.

¹⁶⁵ Rawls ([1993]2005: 137).

¹⁶⁶ *Ibid.*

First, reasonable persons are, when among equals, ‘ready to propose principles and standards as fair terms of cooperation and to abide by them willingly, given the assurance that others will likewise do so’.¹⁶⁷ Second, reasonable persons accept the ‘burdens of judgment’, namely they accept the many ‘hazards involved in the correct (and conscientious) exercise of our powers of reason and judgment in the ordinary course of political life,’¹⁶⁸ which explain how reasonable and rational people will permanently disagree about many matters of value and ethics.

Rawls believes that reasonable persons will accept what he calls the ‘idea of public reason’. The idea of public reason ‘belongs to a conception of a well ordered constitutional democratic society’.¹⁶⁹ A well ordered society is the Rawlsian development of social contract doctrine. It is a society in which:

- (i) everyone willingly accepts and affirms the same principles of justice;
- (ii) the basic structure of society—namely, its basic political, social and economic institutions—is publicly known to satisfy these principles; and
- (iii) citizens have a normally effective sense of justice, that is, they want to do what justice requires of them.¹⁷⁰

The assumption of reasonable persons’ ‘strict compliance’ with principles of justice within a well-ordered society is part of what Rawls calls ‘ideal theory’.

Under these conditions, in a public forum and concerning fundamental political questions,

¹⁶⁷ Rawls ([1993]2005: 49). Note: the moral ‘duty of civility’ also requires that citizens should explain to one another how, at least with regard to constitutional essentials and matters of basic justice, the political positions ‘they advocate and vote for can be supported by the political values of public reason’ (Rawls [1993]2005: 217). Others object that the aims of public reason are best served by permitting debate and deliberation unconstrained by the duty of civility (e.g. Benhabib 2002: 108–112; Bohman 2003; Gaus 2009; Gaus and Vallier 2009; Vallier 2014, 2016). Nothing in my argument will hinge on the civility or otherwise of reasonable persons, so I remain neutral with respect to this feature of Rawls’ theory.

¹⁶⁸ *Ibid*: 55–56.

¹⁶⁹ Rawls (1997: 765).

¹⁷⁰ Rawls (1999b: 4–5, §69; [1993]2005: 35).

‘[a] citizen engages in public reason [...] when he or she deliberates within a framework of what he or she sincerely regards as the most reasonable political conception of justice, a conception that expresses political values that others, as free and equal citizens might also be expected reasonably to endorse’.¹⁷¹

The idea is that reasonable citizens realise they cannot reach agreement on the basis of their irreconcilable comprehensive doctrines. So when fundamental political questions are at stake, Rawls believes that reasonable citizens will appeal only to the kinds of ‘public reasons’ which are acceptable to other reasonable citizens or, in other words, justifiable on the basis of political conceptions of justice.

2.3.2. Political Conceptions of Justice

A political conception of justice has three characteristic features:

- (i) It applies to the ‘basic structure’ of society, namely ‘a society’s main political, social, and economic institutions, and how they fit together into one unified system of social cooperation from one generation to the next’.¹⁷²

Rawls takes the kind of society in question to be ‘a modern constitutional democracy’,¹⁷³ which has an associated ‘criterion of reciprocity’.¹⁷⁴ This is why, as I discuss in the next chapter, Rawls’ criterion of reciprocity acts as a constraint on any political conception of justice.

A political conception of justice has two features which demarcate political relationships in a constitutional democracy from other kinds of relationships. Firstly, because the basic structure is presumptively a ‘structure of basic institutions we enter only by birth and exit only by death’,¹⁷⁵ the political relationship is nonvoluntary, which distinguishes it from

¹⁷¹ Rawls (1997: 773).

¹⁷² Rawls ([1993]2005: 11).

¹⁷³ *Ibid.*

¹⁷⁴ Rawls (1997: 771).

¹⁷⁵ Rawls ([1993]2005: 135-6).

‘associational’ relationships which exist in, for example, religious organisations.¹⁷⁶ Secondly, in a constitutional regime ‘political power is ultimately the power of the public, that is, the power of free and equal citizens as a collective body’.¹⁷⁷ This, Rawls tells us, distinguishes political relationships from ‘the personal and the familial, which are affectional’ relationships.¹⁷⁸ These features have sometimes been taken to denote a sharp public/private distinction in Rawlsian political liberalism.¹⁷⁹ Although I challenge that assumption in the next chapter, one might naturally think, as have feminist critics of Rawls, that political conceptions of justice will have little to say concerning the appropriateness, from a legal perspective, of private sexual activities like robotic sexual violence.

- (ii) A political conception of justice ‘is presented as a freestanding view’,¹⁸⁰ namely it is presented or expounded independently of any comprehensive doctrine.

The notion of presentation is key here: whilst Rawls assumes that ‘all citizens [will] affirm a comprehensive doctrine to which the political conception they accept is in some way related’, a political conception of justice ‘is presented as freestanding and expounded apart from, or without reference to, any such wider background’.¹⁸¹ This gives us another part of Rawls’ answer to the question of political liberalism, namely his account of stability as an overlapping consensus of reasonable comprehensive doctrines—that is, of comprehensive doctrines which ‘accept[...] a constitutional democratic regime and its companion idea of legitimate law’.¹⁸² In an overlapping consensus, stability is achieved when the political conception of justice underlying the legal framework will be endorsed by all reasonable comprehensive doctrines such that the political conception ‘is a module [...] that fits into and can be supported by various

¹⁷⁶ *Ibid.*: 137.

¹⁷⁷ *Ibid.*: 136.

¹⁷⁸ *Ibid.*: 137.

¹⁷⁹ E.g., Okin (1994: 26).

¹⁸⁰ Rawls ([1993]2005: 12).

¹⁸¹ *Ibid.*

¹⁸² Rawls (1997: 766).

reasonable comprehensive doctrines that endure in the society regulated by it'.¹⁸³ The arguments advanced in 2.2 for the legal proscription of robotic sexual violence—depending fundamentally as they did upon Judeo-Christian sexual ethics or an associated account of natural law, the politics of love or a particular moral conception of good character—were, for example, *not* presented as freestanding from 'background' doctrinal commitments.

- (iii) For a political conception of justice, its 'content is expressed in terms of certain fundamental ideas seen as implicit in the public political culture of a democratic society'.¹⁸⁴

This public political culture—comprising 'the political institutions of a constitutional regime and the public traditions of their interpretation'¹⁸⁵—is distinguished from the 'background culture' of civil society—comprising 'comprehensive doctrines of all kinds'.¹⁸⁶ The arguments which I explored in 2.1, ones which supported the legal permissibility of robotic sexual violence, appealed to ideas implicit in public political culture, such as the right to privacy. But the arguments explored in 2.2 for the legal proscription of robotic sexual violence did not: rather, each argument appealed to ideas drawn from the background culture.

The considerations developed in this section illuminate a lacuna in the literature: no one, so far, has advanced an argument for the legal proscription of robotic sexual violence which is acceptable by the standards of public reason. It is to the remedy of this omission that I turn in the next chapter.

¹⁸³ Rawls ([1993]2005: 12).

¹⁸⁴ *Ibid*: 13.

¹⁸⁵ *Ibid*: 13-14.

¹⁸⁶ *Ibid*: 14.

3. Robots and Reciprocity

*In the company of a living enigma man remains alone [...] This subjective game, which can go all the way from vice to mystical ecstasy, is for many a more attractive experience than an authentic relation with a human being.*¹⁸⁷

In this chapter, I articulate and defend an argument which is wholly new for the legal proscription of robotic sexual violence: I contend that there is a public reasons justification for the legal proscription of robotic sexual violence as part of the cultivation of those political virtues which are required by reciprocity in the just society. Recall from the last chapter that public reasons justifications, because they appeal to political conceptions of justice, must have three features: (i) they must establish that the matter is of concern for the basic structure of society; (ii) they must be presented independently of any comprehensive doctrine; and (iii) they must be expressed in terms of ideas which are implicit in the public political culture of a democratic society. Firstly, then, in **3.1**, I explain that private sexual activities are, for Rawls, a matter of concern for the basic structure of society when engagement in them undermines reciprocity; robotic sexual violence, I argue, would undermine reciprocity. My arguments in this section will highlight a novel and suggestive connection between Rawlsian reciprocity and the recent literature on epistemic injustice. Secondly, in **3.2**, I show that my argument for the legal proscription of robotic sexual violence may be presented independently of any comprehensive doctrine because robotic sexual violence cultivates a *political* vice, that is, a character trait which is incompatible with citizenship in a just and stable constitutional regime. Thirdly, I highlight that the ideas underlying my argument for the legal proscription of robotic sexual violence are implicit—and sometimes explicit—in both domestic and international law. **3.3** considers some alternative options for the implementation of this legal proscription in the

¹⁸⁷ Beauvoir (1956: 263).

criminal law and introduces a novel idea that the legal personification of SARs might be a fruitful starting point. It is to the articulation and defence of this idea that I turn in the next chapter.

My arguments in this chapter will draw upon the recent literature on epistemic injustice. There has been a great deal of interest in epistemology regarding the success conditions for testimonial exchanges, that is, what it takes for a hearer to come to know something on the basis of a speaker's say-so.¹⁸⁸ But it is well-known that testimonial exchanges sometimes go systematically wrong. This is because there are problematic patterns in who we tend to believe and in who we tend not to believe, that is, in the *credibility judgments* which underwrite testimonial exchanges. Epistemic injustice occurs when testimonial exchanges systematically go wrong due to facts about a person's social identity (e.g., their race, class or gender). The most widely theorised form of epistemic injustice is called 'testimonial injustice'.¹⁸⁹ Testimonial injustice occurs when someone's assertions are accorded less (or more) credence than they deserve because of prejudice of some kind, such as bias regarding their social identity¹⁹⁰ or harmful 'myths' which render certain explanations of events more likely than others.¹⁹¹ Fricker develops an epistemic virtue ethics, discerning the virtue of testimonial justice and the opposing vice of testimonial injustice.¹⁹² The vice of testimonial injustice is a disposition to commit such acts of epistemic injustice. The virtue of corrective testimonial justice is a disposition to remain aware of, and compensate for, your prejudices by interfering

¹⁸⁸ Some argue that, at a minimum, I need to know what I'm saying (Audi 2006; Burge 1993, 1997; Schmitt 2006). But others reject this requirement (Lackey 2006, 2008; McKinnon 2013, 2015, forthcoming). On other views, I merely need to be reliable in saying true things (Goldberg 2010, 2015). But the dominant view appears to be that I need to be reliable *and* you have to have some reason to think that I am reliable.

¹⁸⁹ The other main variety is 'hermeneutical injustice', which occurs when one is wronged in their capacity as 'a subject of social understanding', that is, when people are denied the conceptual and linguistic resources to make sense of and communicate their experience (Fricker 2007: 7). I focus on the perpetration of testimonial injustice but there is good reason to believe that rape myths also contribute to hermeneutical injustice against victims and survivors of sexual violence (Hänel 2021).

¹⁹⁰ Fricker (2007: 7).

¹⁹¹ Medina (2012, 2013).

¹⁹² Fricker (2003, 2007).

with your estimation of the value of someone's testimony. This corrective virtue, Fricker argues, is cultivated through social training.¹⁹³ A key contribution which I make in this chapter is to show that, in some cases, epistemic injustice is a matter of concern for the basic structure of society and, in those cases, the virtues which characterise epistemically just testimonial exchanges are also political virtues which the state may legitimately seek to promote.

3.1. The Personal and the Political

In this section, I argue that private sexual activities are a matter of concern for the basic structure of society when engagement in them undermines reciprocity (3.1.1). One way to undermine reciprocity is to systematically undermine measures protecting the basic rights of some group of citizens. For example, in 3.1.2, I draw on the recent literature on pornography to argue that pornographic (representations of) sexual violence systematically undermine reciprocity through undermining the ability of women and girls to advance important claims of justice. In 3.1.3, I argue that robotic sexual violence would be a particularly pernicious kind of pornographic sexual violence and, as a result, would further undermine reciprocity.

3.1.1. Justice Behind Closed Doors: A Question of Reciprocity

As we saw in the last chapter, the first hurdle faced by anyone advancing a case for the legal proscription of robotic sexual violence consists in justifying the encroachment of political principles of justice into the private sphere of harmless sexual activities involving only consenting adults. Indeed, it has not always been apparent, especially not to feminist commentators on Rawls, that Rawls' political liberalism is friendly to the application of political principles of justice within the domestic sphere of human life at all. What hope, you might then think, is there for the application of political principles of justice to sex?

¹⁹³ Fricker (2003: 161).

In order to address this concern, I respond to the common misconception that there is a sharp public/private divide implicit in Rawls' insistence that political conceptions of justice apply to the basic structure of society. As I explained in the last chapter, this misconception often derives from the distinction which Rawls draws between those 'political' relationships which characterise the basic structure and those 'associational' and 'affectional' relationships which characterise other institutions, including religious ones and the family. Okin, for example, concludes that 'the distinction Rawls draws between the political and the nonpolitical coincides with his distinction between public and nonpublic'.¹⁹⁴ It would spell trouble for my argument if she were right since, as I explained in the last chapter, paradigmatically 'nonpublic'—i.e., private—activities in constitutional democracies comprise those sexual activities which neither harm nor wrong other people, including, *a fortiori*, robotic sexual violence. But the application of political principles of justice within the private sphere is, for Rawls, much less clear-cut than Okin suggested. On the contrary, political conceptions of justice impose substantive demands of justice on even the most private of human activities including, as I will show in the rest of this section, robotic sexual violence.

In 'The Idea of Public Reason Revisited', Rawls sought to elucidate why political principles of justice constrain even those associational and affectional relationships which characterise the private sphere, arguing forcefully that '[i]f the so-called private sphere is alleged to be a space exempt from justice, then there is no such thing'.¹⁹⁵ This is because Rawls ascribes a certain normative priority to the political relation 'of free and equal citizens [...] within the basic structure of society'.¹⁹⁶ According to Rawls, whilst 'the basic structure alone is the primary subject of justice, the principles of justice still put essential restrictions on the

¹⁹⁴ Okin (1994: 27, 34, 2004: 1556).

¹⁹⁵ Rawls (1997: 791).

¹⁹⁶ Rawls (1997: 769-770, [1993]2005: 12). The priority of the political relation also seems to be conceptual: '[a] domain so-called, or a sphere of life, is not [...] something already given apart from political conceptions of justice [...] but rather is simply the result [...] of how the principles of political justice are applied' (Rawls 1997: 791).

family and all other associations’.¹⁹⁷ Of course, robotic sexual violence is not one of these affectional or associational relationships. But Rawls clarifies that *all* domains of human life are subject to the constraints of the political relation: these ‘principles defining the equal basic liberties and opportunities of citizens always hold in and through all so-called domains’.¹⁹⁸ In particular, Rawls was concerned to emphasise, ‘the equal rights of women [...] are inalienable and protect them *wherever they are*’.¹⁹⁹ The normative priority of the political relation, then, means that there is no private sphere which is exempt from the most basic normative requirements of justice.

You might at this point respond that whilst the private sphere is not wholly exempt from the normative requirements of justice, the normative requirements which apply in our private lives are very basic. In particular, you might think that the normative requirements which are contained essentially in the political relation—independent of any particular political conception of justice—are insufficiently robust to justify anything so ambitious as a legal proscription on robotic sexual violence. Granted, the normative requirements intrinsic in the political relation are basic. But, as I show in this chapter, even these basic requirements will suffice to establish my case.

The normative requirements of the political relation are those contained within the criterion of reciprocity, which acts as ‘the limiting feature’ on all of the ‘many forms of public reason specified by a family of reasonable political conceptions’.²⁰⁰ As Rawls explains, ‘the criterion of reciprocity [...] specif[ies] the nature of the political relation in a constitutional democratic regime as one of civic friendship’.²⁰¹ Reciprocity requires that, when citizens

¹⁹⁷ Rawls (1997: 791).

¹⁹⁸ *Ibid.*

¹⁹⁹ *Ibid.*, italics added.

²⁰⁰ *Ibid.*: 774.

²⁰¹ *Ibid.*: 771. Rawls’ suggestive mention of civic friendship might seem to support the view that part of the moral basis for the idea of public reason is that the practice of public reason is a constitutive part of a valuable

propose terms of cooperation, ‘those proposing them must also think it at least reasonable for others to accept them, as free and equal citizens, and not as dominated or manipulated, or under the pressure of an inferior political or social position’.²⁰² Because of this, Rawls believes that any reasonable political conception of justice must include three features:

R1. ‘a list of certain basic rights, liberties, and opportunities’;²⁰³

R2. ‘an assignment of special priority to those rights, liberties, and opportunities, especially with respect to the claims of the general good and perfectionist values’;²⁰⁴ and

R3. ‘measures ensuring for all citizens adequate all-purpose means to make effective use of their freedoms’.²⁰⁵

This is because reciprocity imposes substantive demands upon the kinds of principles which free and equal citizens may reasonably be expected to accept when determining principles for a constitutional democracy: ‘[f]or what reasons can both satisfy the criterion of reciprocity and justify denying to some person religious liberty, holding others as slaves, imposing a property qualification on the right to vote, or denying the right of suffrage to women?’.²⁰⁶

It follows from my discussion so far that political principles of justice apply to private sexual activities when engagement in those activities would undermine the requirements which reciprocity imposes always and everywhere upon free and equal citizens. As I explained in the last chapter, robotic sexual violence cases like Robotic Rape would not seem to imply the violation of any person’s basic rights, liberties or opportunities. Nor would robotic sexual violence cases seem to undermine the special priority which reciprocity assigns to those basic

relationship (Ebels-Duggan 2010; Leland 2019; Leland and van Wietmarschen 2017; Lister 2013). However, there is reason to doubt that civic friendship alone can ground the idea of public reason (Billingham and Taylor 2023).

²⁰² Rawls (1997: 770).

²⁰³ *Ibid*: 774.

²⁰⁴ *Ibid*.

²⁰⁵ *Ibid*.

²⁰⁶ *Ibid*: 771.

rights, liberties or opportunities. My contention in this section is rather that robotic sexual violence cases would undermine measures which aim to ensure adequate all-purpose means for women and girls to make effective use of their freedoms. In order to establish the truth of this contention, it will be instructive to see how existing forms of pornography undermine reciprocity in a similar way.

3.1.2. Beyond the Porn Wars: Rape Myths and the Reporting Crisis

During the ‘porn wars’ of the 1970s and 1980s, so-called ‘anti-porn’ feminists sought to establish that pornography—namely, in broad terms, ‘sexually explicit material [...] primarily designed to produce sexual arousal in viewers’²⁰⁷—either constitutes or causally contributes to the perpetration of sexual violence against women and girls.²⁰⁸ Of course, one may undermine reciprocity through undermining or deprioritising the right of all persons to be secure from violence, including sexual violence (contravening R1 and R2). But I take no stand in Part One of this dissertation as to whether the consumption of pornography undermines or deprioritises the basic rights of women and girls in this regard. Rather, I wish to develop a new line of argument concerning the relationship between pornography and sexual violence, one which focuses instead on how certain kinds of pornography undermine the ability of women and girls to successfully advance important claims of justice.

Besides undermining R1 or R2, another way to contravene the requirements of reciprocity is to undermine R3, that is, those measures which purport to ensure ‘adequate all-purpose means’ for citizens to make ‘effective use’ of their basic rights to be secure from sexual violence. These measures include the provision of adequate enforcement mechanisms to ensure that redress is effectively available to all persons in the eventuality that these basic rights are

²⁰⁷ West (2022).

²⁰⁸ On equality and, relatedly, the elimination of sexual violence as a reason for pornography regulation: MacKinnon (1993).

violated. To undermine reciprocity, then, it suffices to systematically undermine those enforcement mechanisms, rendering them wholly inadequate for their purpose. These enforcement mechanisms are famously already in a dire state when it comes to the enforcement of the rights of women and girls to be secure from sexual violence.²⁰⁹ One problem is that crimes are often not reported.²¹⁰ The Women's Aid Federation of England and Wales (WAFEW) notes that one of the key obstacles to reporting is the existence of 'pervasive rape myths within society which shape responses to survivors'.²¹¹ Another problem is that even when crimes are reported, the credibility of survivor testimony is often wrongfully diminished in court by appeal to rape myths. As WAFEW documents: 'Myths and stereotypes are [...] perpetuated in the courtroom, and are inconsistently challenged by prosecutors, for example how a 'true' victim will behave in the aftermath of a rape'.²¹² In addition, these rape myths interact with sexual history evidence concerning the victim which 'attempts to imply to the courtroom that the survivor is [...] less credible as a witness'.²¹³ More generally, research shows that rape myth acceptance directly correlates with a tendency to dismiss allegations.²¹⁴

One motor for the propagation of rape myths in society is the consumption of pornographic sexual violence—viz., pornographic representations of sexual violence for the user's sexual gratification. It has been argued since the 1985 Meese Report that, among other things, widespread consumption of pornographic sexual violence in a society makes individuals more ready to accept rape myths and desensitises them to sexualised violence.²¹⁵ Research in the psychology and sociology of sexual violence supports this claim: front-line

²⁰⁹ The UK Parliament Home Affairs Committee (2022: ¶43) reports that 'victims and survivors of rape are being failed by our criminal justice system' due to 'unacceptable' and 'exceptionally low volumes of rape charges and prosecutions'.

²¹⁰ Fewer than one in six (16%) female victims and fewer than one in five (19%) male victims aged 16 to 59 years of sexual assault by rape or penetration since the age of 16 years reported it to the police (ONS 2021).

²¹¹ WAFEW (2021: ¶4).

²¹² WAFEW (2021: ¶12). See also Smith and Skinner (2012, 2017).

²¹³ WAFEW (2021: ¶12). See also McGlynn (2017).

²¹⁴ Krahé, Temkin, and Bieneck (2007).

²¹⁵ Dwyer (1995: legal appendix); Einsiedel (1992: 266); Langton (1993: 310).

staff working with sex offenders and survivors consistently document how widespread exposure to and consumption of pornographic sexual violence plays a significant part in shaping attitudes and behaviours towards women and girls which make sexual violence seem like a normal part of sex.²¹⁶ Recent research more generally shows that higher pornographic sexual violence incidence in a society changes sexual scripts to normalise sexualised violence towards women and girls. The theory of sexual scripts has its origins in Gagnon and Simon²¹⁷ and, broadly, concerns the norms governing sexual behaviours in a social context. For example, research suggests that exposure to and consumption of pornographic sexual violence are associated with the normalisation of high-risk and violent sexual behaviours,²¹⁸ objectifying attitudes towards women,²¹⁹ physical aggression²²⁰ and discriminatory attitudes, especially against Asian women.²²¹ As a result, Vera-Gray and others argue that ‘sexual violence is a normative sexual script in mainstream online pornography, with significant implications at a social level for understandings of the difference between sexual pleasure and sexual harm’.²²²

It bears emphasis that my claim here is *not* that representations of sexual violence in mainstream pornography increase the risk or incidence of actual sexual violence, an issue on which this dissertation remains neutral.²²³ Rather, my claim is that pornographic sexual violence produces and reproduces rape myths. Rape myths, I submit, produce and reproduce epistemic injustice against women and girls because they reinforce problematic patterns in the credibility judgments which underwrite testimonial exchanges with victims and survivors of

²¹⁶ See, for example, Wyre (1992: 241-2), documenting psychological evaluation of sex offenders, and the recent report for the British Government Equalities Office which found that ‘All [...] Frontline workers [...] acknowledged [...] pornography as an influential factor for harmful sexual behaviours towards women and girls’ in the sense that it ‘contributes to a conducive context facilitating these behaviours’ (Upton *et al.* 2021: 8).

²¹⁷ Gagnon and Simon (1973).

²¹⁸ Braithwaite *et al.* (2015).

²¹⁹ Sun *et al.* (2016); Fritz and Paul (2017).

²²⁰ Shor and Seida (2019, 2020).

²²¹ Zhou and Paul (2016).

²²² Vera-Gray *et al.* (2021: 1244).

²²³ Thanks to an anonymous reviewer for pressing me on this point.

sexual violence. For instance, pornographic sexual violence may make it seem more likely that a victim or survivor consented to a given sexual act through depicting that kind of act as a normal, enjoyable or consensual part of sex. Indeed, we have seen that there appears to be a widespread and wrongful deflation of the credibility of women's and girls' testimony about sexual violence, attributable in part to rape myths which render certain explanations of events more salient than others to participants in a testimonial exchange about sexual violence.

It is instructive to analyse the problem with rape myths in terms of Haslanger's distinction between the 'manifest' concept of a term—that is, its explicit official and formal definition—and its 'operative' concept—that is, the implicit definition that would be extrapolated from its actual usage in a given community.²²⁴ Rape myths, then, divorce the manifest concept (the legal definition of 'rape') from its operative concept (the actual usage of 'rape' in a given community). As we have seen, this divorce contributes to the widespread and wrongful deflation of the credibility of women's and girls' testimony about sexual violence. This is a serious testimonial injustice. It undermines reciprocity because it undermines enforcement mechanisms designed to ensure 'adequate all-purpose means' for citizens to make 'effective use' of their basic rights to be secure from sexual violence. Indeed, the systematic failure to enforce the rights of women and girls to be secure from sexual violence means that women and girls as a social group face wholly inadequate means to make effective use of their basic freedoms. As Green puts it:

'Women who have not been, and may never be, victims of rape nonetheless modify their lives for fear of it, or because their parents fear it: they feel obliged to avoid certain places, to be careful going out at certain times, to give thought to how their dress and

²²⁴ Haslanger (2005: 14).

manner will be received—in general, to remain unwillingly vigilant. Few men have experiences that are remotely similar.’²²⁵

And rape myths, propagated at least in part by consumption of and exposure to pornographic sexual violence, are known to be part of the problem. What does all this have to do with the legal proscription of robotic sexual violence?

3.1.3. Autonomous Pornography

It is standard in the SARs literature to draw an analogy between pornography and SARs.²²⁶ Indeed, by all accounts, SARs may be viewed as a new kind of pornography. It is easy to see why this is the case if pornography is, as mentioned above, ‘sexually explicit material [...] that is primarily designed to produce sexual arousal in viewers’,²²⁷ ‘any object that has been manufactured to satisfy sexual desire through its sexual consumption or other sexual use’²²⁸ or, more emotively, ‘sex[ual activity] between people and things, human beings and pieces of paper, real men and unreal women’.²²⁹ Of course, the ‘material’, ‘object’ or ‘thing’ in question is traditionally assumed to be verbal, pictorial or videographic; but why not also robotic? This presumption would already give us good reason, based upon the arguments advanced in the previous section, to think that robotic sexual violence would be a matter of concern for the basic structure of a just society.²³⁰ But for reasons which I develop below, robotic sexual

²²⁵ Green (2013: 491).

²²⁶ See, for example, Danaher (2017a, 2017b); Sparrow (2017); Eskens (2017); Gersen (2019).

²²⁷ West (2022).

²²⁸ Vadas (2005: 177).

²²⁹ MacKinnon (1993: 109). Elsewhere, MacKinnon describes pornography somewhat differently, as ‘the graphic sexually explicit subordination of women through pictures and words’ (MacKinnon 1987: 176). In that case, we would need to expand the concept to include the subordination of women ‘through pictures and words and autonomous robots’. But this seems true to the spirit of MacKinnon’s account.

²³⁰ This line of argument has not been developed in the literature but is suggestive of how a Rawlsian political liberal might justify the UK’s somewhat controversial criminalisation of possession of an ‘extreme pornographic image’, which includes some if not all depictions of rape (Criminal Justice and Immigration Act 2008, Pt.5). For criticism of the Act: Carline (2011); Cowen (2015).

violence would be an especially pernicious form of pornographic sexual violence from the perspective of rape myth propagation.

Naively, this might seem unlikely because of the ways in which SARs differ from existing kinds of pornography. As I explained in chapter one, SARs are social robots which, unlike existing pornography, look, speak and act rather like human sexual partners. The experience of enacting robotic sexual violence might, then, be more lifelike. By themselves lifelike representations of sexual violence when only consenting adults are involved are not a matter of concern for the basic structure. As Gersen emphasises, '[i]t is hardly uncommon for human partners to engage together in various degrees of consensual role-play that includes fantasies of force, coercion, roughness, resistance, domination, and submission'.²³¹ Unlike pornographic sexual violence, which propagates rape myths by promulgating ready-made sexual scripts and stereotypes through verbal, pictorial and videographic means, robotic sexual violence might at first blush seem to offer more scope for human imagination, creating a more private experience in which the SAR's user may experiment with different kinds of sexual fantasy. In other words, the *autonomous* nature of SARs might initially seem to open up new possibilities for those experiments in living which are already a legitimate part of human sexual relationships.²³²

In response, I invite you to recall the distinction drawn in chapter one between general and narrow AI. If SARs were powered by general AI—that is, if they displayed the kind of sophistication, flexibility and generality in their intelligence which human beings do—the objection considered here might have some bite: in particular, if the autonomous nature of SARs were to involve the kind of creative or quirky autonomy which we see in adult human beings, then robotic sexual violence might not be the kind of motor for propagating harmful

²³¹ Gersen (2019: 1795). C.f. Crépault and Couture (1980); Critelli and Bivona (2008).

²³² For notes of optimism in this vein: Levy (2007).

stereotypes and rape myths that pornographic sexual violence seems to be.²³³ But SARs will not be powered by general AI for the foreseeable future. Rather, SARs will involve narrow AI endowed, as I explained in **1.2.3**, with at least some bottom-up machine learning capabilities to enable them to respond like a human sexual partner to new and unexpected situations. And, the AI ethics literature reveals, nothing propagates harmful stereotypes and bias quite like bottom-up machine learning. One problem is that bottom-up machine learning bases the AI system's future behaviour on regressions taken over potentially biased input data and therefore incorporates and augments existing biases.²³⁴ As we've seen, the manifest and operative concepts of sexual violence are radically divorced in much of our sexual culture, a problem which is often colloquially referred to as 'rape culture'. It is therefore difficult to see how we could find sufficient unbiased training data to ensure that SARs would not learn—and learn with unprecedented efficiency—how to systematically reproduce rape culture in their sexual behaviours.²³⁵ Another problem is that bottom-up machine learning is very difficult to audit because it inherently lacks transparency²³⁶: some of the best machine learning techniques (known as 'random forests' machine learning) result in wholly inscrutable 'black-box' AIs, which threaten the insinuation of rape myths and harmful stereotypes (derived from the sexual scenarios of its training data) into the SAR's behaviour.²³⁷ And, unlike in the ready-made scenarios of existing pornography, the problematic biases ingrained in SARs are not even in principle scrutable to human users.²³⁸

²³³ SARs powered by general AI would raise other issues, which I touch on in chapter five.

²³⁴ Cohen (m.s.); Feldman *et al.* (2015); Forrest (2021); Kirkpatrick (2017). C.f. Sinnott-Armstrong and Skorburg (2021).

²³⁵ There is a parallel to be drawn here between the effects of rape culture on *narrow* as compared to *general* AI and its effects on *minors* as compared to *adults*. Without an adequate level of autonomy, it appears that minors are much more susceptible than adults to uncritically reproducing sexual scripts and rape myths from mainstream pornography in their own sexual behaviours (Adarsh and Sahoo 2023).

²³⁶ Brennan and Oliver (2013: 558).

²³⁷ Berk and Bleich (2013).

²³⁸ In limited cases, there may be ways to audit certain kinds of black-box AI for bias: see Adler *et al.* (2016).

SARs, then, would be a new kind of autonomous pornography. The nature of their autonomy means that robotic sexual violence would be a much more effective mechanism for propagating harmful stereotypes and rape myths than existing kinds of pornographic sexual violence. This mechanism undermines reciprocity in the just society and is, therefore, a legitimate concern for political principles of justice applied to the basic structure.²³⁹ This brings me to the second concern for the application of political principles of justice to the regulation of robotic sexual violence: can arguments for the legal proscription of robotic sexual violence be presented independently of any comprehensive doctrine?

3.2. Virtues of the Polity

In this section, I argue that, in addition to being a matter of legitimate concern for the basic structure, a legal proscription on robotic sexual violence may be defended without appeal to any comprehensive doctrines and based upon ideas implicit in public political culture. First, I articulate and defend the Rawlsian distinction between moral virtues and vices and political ones (3.2.1). Then, I show that in contexts where the perpetration of testimonial injustice undermines reciprocity, testimonial injustice there becomes a political vice (3.2.2). 3.2.3 argues that legal measures cultivating the corrective political virtue of testimonial justice may be justified by appeal to ideas already implicit in public political cultures. As a result, the legal proscription of robotic sexual violence is part of reasonable measures to promote reciprocity, which is a requirement of all political conceptions of justice.

3.2.1. When is virtue political?

At this point in my argument, one might pertinently point out that lots of activities undermine (or sustain) reciprocity but do not thereby warrant legal proscription (or prescription). Your

²³⁹ The arguments advanced so far shine light on a novel public reasons argument for the legal proscription of AI-generated pornography, as I discuss in chapter four. But I will suggest that the more humanlike function of SARs may prompt us to consider a rather distinctive legal framework for their regulation.

smoking cigarettes, for example, might set a bad example to others, who will thereby associate this habit with high intelligence and academic success; in this way, your smoking might undermine measures (such as anti-smoking campaigns) designed to ensure for all citizens ‘adequate all-purpose means to make effective use of their freedoms’ (assuming that many freedoms cannot be so easily enjoyed by, for example, cancer patients). Nonetheless, it seems inappropriate for the state to ban smoking by appeal to the viciousness of setting a bad example. On the flip-side, daily Buddhist meditation practices for police officers might cultivate certain forms of compassionate attentiveness, which would sustain reciprocity by combating the testimonial injustices of rape myths. But even if this link could be empirically proven, it would still seem inappropriate for the state to legally prescribe daily Buddhist meditation practices by appeal to the virtue of compassionate attentiveness. So, the objection goes, there are certain dispositions the cultivation or suppression of which is not an appropriate exercise of state power, irrespective of their effects on reciprocity.

This objection turns on a misunderstanding of the nature of public justification. The examples given above serve to highlight not the irrelevance of reciprocity in public justification, but rather that the reasons which we appeal to in public justifications matter too. It is true that it would be inappropriate for free and equal citizens to seek to justify exercises of political power by appeal to the viciousness of setting a bad example or the virtue of compassionate attentiveness. This is because the justification would be *presented* in terms of ‘moral virtues’, that is, those virtues which are appropriate ‘to the relations between individuals’, which hold in, for example, religious and familial circles.²⁴⁰ But it would *not* be inappropriate for free and equal citizens to seek to justify exercises of political power by appeal to the effects of certain virtues and vices on reciprocity. This is because the justification would be presented in terms of ‘political virtues’, that is, those virtues which are intrinsic to the

²⁴⁰ Rawls ([1993]2005: 195).

political relation because they ‘characterise the ideal of a good citizen in a democratic state’.²⁴¹ The categories overlap: as Rawls explains, ‘political virtues are identified and justified by the need for certain qualities of character in the citizens of a just and stable constitutional regime’.²⁴² But ‘[t]his is not to deny that these same characteristics, or similar ones, might not also be nonpolitical virtues insofar as they are valued for other reasons within various particular comprehensive views.’²⁴³ This is why, in the objection under consideration, the moral vice of setting a bad example and the moral virtue of compassionate attentiveness would also be matters of political vice and virtue provided that the empirics of their effects on reciprocity could be corroborated. This prompts the following question: can we justify the legal proscription of robotic sexual violence by appeal only to political virtues or vices?

3.2.2. Testimonial Justice as a Political Virtue

The disposition to commit testimonial injustice against sexual violence survivors, I have argued, undermines the political relation; it is, therefore, a political vice. The corrective political virtue is, as Fricker explains, the disposition to remain aware of and compensate for your prejudices, a virtue acquired through social training.²⁴⁴ Now, the cultivation of political virtues is a legitimate exercise of political power, for Rawls, because that cultivation may be justified independently of comprehensive doctrines. As Rawls explains:

‘[I]f a constitutional regime takes certain steps to strengthen the [political] virtues of toleration and mutual trust, say by discouraging various kinds of religious and racial discrimination (in ways consistent with liberty of conscience and freedom of speech) it does not thereby become a perfectionist state. Rather, it is taking reasonable measures

²⁴¹ *Ibid*: 194-5.

²⁴² *Ibid*: 194, n.29.

²⁴³ *Ibid*.

²⁴⁴ Fricker (2003: 161).

to strengthen the forms of thought and feeling that sustain fair social cooperation between its citizens regarded as free and equal'.²⁴⁵

My first contention, then, is that a constitutional regime may by the same logic take certain steps to strengthen the political virtue of testimonial justice, by discouraging certain kinds of what may appropriately be called *gender* discrimination, including the testimonial injustices faced by sexual violence survivors.²⁴⁶ I discuss this notion of gender discrimination further in

3.2.3.

I argued in 3.1.3 that robotic sexual violence would contribute to the development of the disposition to commit testimonial injustice against sexual violence survivors. So robotic sexual violence would contribute to the development of a political vice. Steps to foster the corrective political virtue of testimonial justice, then, appear to be part of 'taking reasonable measures to strengthen the forms of thought and feeling that sustain fair social cooperation between its citizens regarded as free and equal'. These steps, because their justification appeals to matters of concern for the basic structure and is presented without appeal to any comprehensive doctrine, are not incompatible with political conceptions of justice. My second contention in this section, therefore, is that one such step appropriately involves the legal proscription of robotic sexual violence. In order to assess this latter contention, it will help to explore to what extent the ideas underlying my proposal are implicit in public political culture.

3.2.3. Conventions Concerning Discrimination

As Rawls explains, '[i]n a democratic society there is a tradition of democratic thought, the content of which is at least familiar and intelligible to the educated common sense of citizens

²⁴⁵ Rawls ([1993]2005: 195).

²⁴⁶ It goes without saying that 'sexual violence survivors' includes people of all genders. However, sexual violence remains a gendered issue given its disproportionate impact on women and girls (UN Secretary-General 2025; UN Women 2025; UNICEF 2025).

generally’,²⁴⁷ which should be ‘seen as a fund of implicitly shared ideas and principles’ which may be used in public reasons justifications. In order to show that the legal proscription of robotic sexual violence which I defend in this chapter is indeed justified by appeal to public reasons, then, I need to show that its content is expressed in terms of certain fundamental ideas seen as implicit in the public political culture of a democratic society. In this subsection, I show that the idea of cultivating the political virtues associated with combating epistemic injustice are implicit within public political culture, both domestically and internationally.

Within the public political culture of the UK, it is encouraging to see that the dangers which rape myths and stereotypes pose to the success of testimonial exchanges are recognised as a serious legal issue. The Crown Prosecution Service’s legal guidance for sexual offences devotes a whole chapter to ‘[t]ackling rape myths and stereotypes’, insisting that

‘Rape myths and stereotypes [...] need to be identified and addressed where they arise in order to ensure a proper case-strategy and effective advocacy when presenting a case at trial’.²⁴⁸

Moreover, the guidance recognises that

‘technological advances’ and associated ‘changes in sexual behaviours’ in recent years mean that ‘there is an increased risk that myths and stereotypes may arise in rape cases in relation to the complainant, the offender, consent issues and/or the circumstances of the alleged offending’.²⁴⁹

The guidance also cites *Miller v R*,²⁵⁰ where the Court of Appeal endorsed the use of judicial directions to counter the risk of rape myths and stereotypes wrongfully deflating the credibility

²⁴⁷ Rawls ([1993]2005: 14).

²⁴⁸ CPS (2021: ch.4).

²⁴⁹ *Ibid.*

²⁵⁰ [2010] EWCA Crim 1578.

of survivor testimony. In other words, judges should direct counsel and juries to combat the testimonial injustices faced by sexual violence survivors, especially in light of how technological advances have affected sexual norms.

But at the international level, something like a call to cultivate political virtues to ensure adequate all-purpose access for women and girls to make effective use of their basic freedoms has been an explicit requirement under international law since the 1979 Convention on the Elimination of all forms of Discrimination Against Women (CEDAW), which has been ratified by 189 countries.²⁵¹ Indeed, CEDAW explicitly understands eliminating discrimination against women in terms of guaranteeing the ‘recognition, enjoyment or exercise by women [...] of human rights and fundamental freedoms in the political, economic, social, cultural, civil or any other field’.²⁵² But CEDAW was and remains a radical document, calling on state parties to

‘take all appropriate measures: [...] [t]o modify the social and cultural patterns of conduct of men and women, with a view to achieving the elimination of prejudices and customary and all other practices which are based on the idea of the inferiority or the superiority of either of the sexes or on stereotyped roles for men and women’.²⁵³

Prejudice and stereotypes against women, CEDAW indicates, are not only personal vices but matters of international public political concern. Their elimination has therefore been a requirement under international law for several decades. One appropriate measure to achieve this end, which I have defended by appeal to public reasons acceptable to all reasonable citizens, is the legal proscription of robotic sexual violence. How, though, should this legal proscription be implemented?

²⁵¹ Note: the USA has merely signed it and is, therefore, not legally bound by CEDAW. Still, ‘[t]he signature [...] creates an obligation to refrain, in good faith, from acts that would defeat the object and the purpose of the treaty’ (Vienna Convention on the Law of Treaties 1969, Arts.10, 18).

²⁵² CEDAW, Art.1.

²⁵³ CEDAW 1979, Art.5.

3.3. Options for implementation

There appear to be several ways in which we might implement the legal proscription of robotic sexual violence. One way is to create specific criminal law prohibitions, regulations and safeguards applicable to the production, sale and use of SARs. We should, for instance, ban those SARs which are specifically designed for the purposes of enacting robotic sexual violence.²⁵⁴ However, as chapter one explained, such a ban will not be enough. This is because even a tried-and-tested SAR which reliably consents enthusiastically to all sexual advances may withhold consent on any new occasion. This unpredictability is a feature of the bottom-up machine learning which autonomous AI systems, like SARs, incorporate. So, the legal proscription of robotic sexual violence calls for more targeted steps. For instance, we need to impose a duty on the manufacturers of SARs to ensure that adequate safeguards are imposed top-down to prevent the misuse of SARs as well as a duty on users not to engage in robotic sexual violence.²⁵⁵ There might be certain practical difficulties for the enforcement of these duties, particularly—given privacy considerations—those which apply to the users of SARs. However, as I discuss in the next chapter, there are expressive and educative functions of the law which apply even in cases where it may, for one reason or another, be only rarely coercively enforced.

Another approach would be to legally proscribe the production, sale and use of SARs altogether. One benefit would in principle be the effective prevention of robotic sexual violence. I do not wish to rule out the proposal. But there are two reasons why I find it unsatisfying. Firstly, many believe that, at least as a general rule, something should be morally

²⁵⁴ C.f. Danaher's (2017a) argument for the legal proscription of 'rape and child sexual abuse robots'. Danaher appears to be referring to sex dolls rather than SARs, because the robots in question are not autonomous.

²⁵⁵ Of course, the kinds of penalties imposed in the event of a breach of these duties would be significantly less severe than the kinds of penalties applicable for the commission of sexual violence against a human being.

wrong in order to warrant criminalisation.²⁵⁶ On this view, criminalising the production, sale and use of SARs appears overinclusive. Indeed, I have neither seen nor defended a compelling public reasons argument against the production, sale and use of SARs *per se*. Of course, some are happy to accept that the criminal law may be overinclusive. Classic examples of allegedly overinclusive crimes include statutory rape and drunk driving offences, the claim being that there are isolated cases of each offence which are not morally wrong.²⁵⁷ Because of this, Green argues that moral wrongness is not a precondition for criminalisation. But even if it is true that some crimes are overinclusive as Green suggests, this does not mean that we should embrace overinclusivity as a perennial feature of the criminal law. On the contrary, overcriminalisation would seem to undermine the message of condemnation which criminalisation intuitively conveys. Indeed, Coffee suggests that ‘stigma is a scarce resource’ which diminishes in expressive power when overused.²⁵⁸ Because of this, I believe that we should be cautious about criminalising the production, sale and use of SARs altogether.

I do not purport to have offered any conclusive arguments against either of the two foregoing approaches. However, I will argue in the next chapter that there is something rather distinctive about SARs which might motivate a distinctive approach to their legal regulation. In essence, my proposal will be that we should implement the legal proscription of robotic sexual violence by legally personifying SARs and ascribing to them at least one legal right, namely the right to be secure from robotic sexual violence. That right corresponds to the correlative legal duties adverted to above which apply to the producers of SARs as well as to

²⁵⁶ This is what Duff (2014: 218) calls ‘negative legal moralism’. It is endorsed by, e.g., Duff (2014: 219); Feinberg (1984); Husak (2007: ch.2); Lamond (2007: 625). There is a well-known objection to negative legal moralism, which appeals to the distinction between two classes of crimes: ones which are *mala in se*—viz., ‘naturally evil as adjudged by the sense of a civilized community’, such as murder—and ones which are *mala prohibita*—viz., ‘wrong only because a statute makes it so’, such as speeding (*State v. Horton* [1905] 51 S.E. 945). But Duff (2018) believes that *mala prohibita* are compatible with the claim that conduct must be morally wrong in order to warrant criminalisation, even if that conduct need not be morally wrong prior to its legal proscription. For criticisms: Tadros (2016); Green (2020).

²⁵⁷ I am sceptical but see Green (2020).

²⁵⁸ Coffee (1991: 238).

their users. However, the ascription of legal personhood, I will suggest, involves something over-and-above the ascription of rights and duties, something which better engages with the distinctive role which SARs are designed to play in human relationships. One way to get a flavour of what is to come appeals to Stone's argument in the environmental ethics literature that the legal personification of certain natural objects accomplishes something over-and-above mere legal proscriptions against their misuse.²⁵⁹ In part, this is due to the way in which language itself influences legal thinking:

‘Introducing the notion of something having a ‘right’ (simply *speaking* that way), brings into the legal system a flexibility and open-endedness that no series of specifically stated legal rules [...] can capture [...] [such that] a society in which it is stated, however vaguely, that ‘rivers have legal rights’ would evolve a different legal system than one which did not employ that expression, even if the two of them had, at the start, the very same ‘legal rules’ in other respects.’²⁶⁰

These considerations are, at this stage, purely suggestive but I turn, in the next chapter, to the full articulation and defence of the first argument for the legal personhood of SARs.

3.4. Concluding Remarks

In this chapter, I articulated and defended a public reasons basis for the legal proscription of robotic sexual violence. This is because robotic sexual violence would undermine reciprocity and is, therefore, a matter of concern for the basic structure of society (3.1). In addition, the legal proscription of robotic sexual violence may be presented independently of any comprehensive doctrine and by appeal to ideas implicit in public political culture: robotic sexual violence cultivates a particular political vice, which is condemned both domestically (in

²⁵⁹ Stone (2010).

²⁶⁰ *Ibid*: 47.

basic English legal institutions) and internationally (via CEDAW's widespread ratification) (3.2). 3.3 explored some alternative options for implementing the legal proscription of robotic sexual violence. In the next chapter, I articulate and defend a novel approach, one which attends to exactly what it means to be a legal person.

4. Robots and Persons

*If you prick us do we not bleed? If you tickle us do we not laugh? If you poison us do we not die? And if you wrong us shall we not revenge?*²⁶¹

In the last chapter, I showed that there is a public reasons justification for the legal proscription of robotic sexual violence. To complete my response to the robotic sexual violence problem, in this chapter, I sketch the contours of a legal framework for the implementation of that legal proscription. My contention is that the ascription of legal rights to SARs to be secure from robotic sexual violence would be part of ‘taking reasonable measures to strengthen the forms of thought and feeling that sustain fair cooperation between [...] citizens regarded as free and equal’.²⁶² An important corollary of this claim is that SARs are the kinds of entities which may appropriately bear legal rights. In other words, SARs should be ascribed legal personhood. In **4.1**, I argue for a ‘functionalist’ interpretation of legal personhood, one on which ‘a person is an agent who can perform effectively in the space of [legal] obligations’.²⁶³ According to this functionalist interpretation of legal personhood, I show in **4.2** that SARs should be ascribed legal personhood, including legal rights against robotic sexual violence. I conclude in **4.3** that the ascription of legal personhood to at least some autonomous robots would be a legally efficacious and morally appropriate way to reflect our evolving relationships with them.

4.1. The Grounds of Legal Personhood

It is often assumed that SARs would not be the kinds of entities which may appropriately bear legal rights because they do not possess the kinds of properties which are generally believed to

²⁶¹ These lines are spoken by Shylock, asserting his personhood, in Shakespeare’s *The Merchant of Venice* III(i).

²⁶² Rawls ([1993]2005: 195).

²⁶³ List and Pettit (2011: 173). List and Pettit (2011) use the term ‘performative’ instead of ‘functionalist’. But I prefer to leave the connection between the account of personhood and the philosophy of mind doctrine of functionalism explicit.

underpin intrinsic moral status.²⁶⁴ Indeed, there is remarkable agreement that SARs ‘do not meet the criteria for legal or moral personhood’²⁶⁵ because they ‘are ‘mere’ machines’,²⁶⁶ that is, they are not ‘sentient being[s]’,²⁶⁷ are ‘[in]capable of thinking and feeling’²⁶⁸ or, as one Taoist analysis has it, have ‘no qi-energy’.²⁶⁹ One exception to this concord appears in Frank and Nyholm. In their closing remarks, the authors gesture towards the possibility that SARs might be ‘brought into the legal community’, suggesting that we ‘take[...] seriously’ the extension of ‘a form of ‘electronic personality’” to SARs.²⁷⁰ But no elaboration of this proposal is forthcoming in the SARs literature. My contribution in this chapter will be to offer principles to remedy this omission. To begin with, I suggest that legal personhood does necessarily depend upon the possession of specific intrinsic properties (4.1.1). Rather, ascriptions of legal personhood should be based upon a constructive interpretation of law, one which both ‘fits’ and ‘justifies’ the extensional facts about actual ascriptions of legal personhood (4.1.2). Functionalism about legal personhood satisfies these dual constraints (4.1.3). In the next section, I apply functionalism to the question of whether SARs should be legal persons.

4.1.1. Personhood Legalism and Personhood Moralism

According to one school of thought, which I call ‘personhood legalism’ or ‘legalism’ for short, the assumption in the SARs literature that legal personhood is a moral matter at all might appear utterly mistaken. For instance, in their defence of the legal personhood of certain kinds of AI systems, Chopra and White suggest that ‘[l]egal entities are recognized as such in order to facilitate the working of the law in consonance with social realities’.²⁷¹ Indeed, legalists take

²⁶⁴ C.f., Darling (2014: 1, n.4).

²⁶⁵ Danaher (2017b: 73).

²⁶⁶ Sterri and Earp (2021).

²⁶⁷ Jecker (2020: 90).

²⁶⁸ Sparrow (2017: 467).

²⁶⁹ Zhang (2021: 90).

²⁷⁰ Frank and Nyholm (2017: 322).

²⁷¹ Chopra and White (2011: 155). C.f., Turner (2019).

ascriptions of legal personhood to be derived only from other legal norms. The paradigm legalist is Kelsen, whose theory of legal personhood has three key aspects. First, being a legal person is identical to having legal rights and obligations.²⁷² Second, since the law, for Kelsen, is an order of norms that prescribe the behaviour of natural persons²⁷³ and legal persons are simply bundles of legal norms, natural persons—whose behaviour is governed by legal norms—are legal persons.²⁷⁴ Third, Kelsen rejects any principled distinction between natural persons and artificial persons because the behaviour of so-called artificial persons, like corporations, encodes the behaviour of natural persons and is, therefore, the subject of legal norms. As a result, natural and artificial persons are both simply legal persons, that is, bundles of rights and obligations,²⁷⁵ which ultimately address the actions of individual human beings.²⁷⁶ The refusal to ground legal principles in underlying moral principles is definitional of Kelsen's 'pure' theory of law, which 'characterizes itself as a 'pure' theory of law because it aims at cognition focused on the law alone', that is, the derivation of legal principles based only on other legal principles, independent of moral, political or scientific considerations.²⁷⁷ As a result, legalists of the Kelsenian tradition would be unmoved by arguments against ascribing legal personhood to SARs on the basis that SARs lack intrinsic moral status.

There appear to be two principal problems with legalism, one practical and the other more principled. First, the practical problem. There is an attractive principle of adjudication, famously developed in Dworkin's *Law's Empire*, on which 'legal reasoning is an exercise in

²⁷² Kelsen (1925: 63-4). See also: Kelsen (1992: 46-7, 1945: 93-5, 2005: 172-3, 2015: 6); c.f. Altwick (2015: 236-9). Kelsen compares the separation of the legal person from its rights and obligations to the separation of a tree from its trunk, branches and blossoms: to do so would be impossible (2015: 172, c.f. 1925: 62). As Kirste puts it, Kelsen explains the legal person 'by making it disappear' (2000: 33).

²⁷³ Kelsen (2005: 30-2, 1991: 28-9).

²⁷⁴ Still, Kelsen grants that legal norms that address non-human beings are possible in theory. Indeed, he explores the practice of animal trials arising from animist beliefs and the possibility of training animals (Kelsen 1925: 62-3, 1923: 518-9, 1945: 3-4, 2005: 32-3, 1991: 89-91, 1992: 29).

²⁷⁵ Kelsen (2005: 174).

²⁷⁶ *Ibid*: 171-178, 190-1. On so-called 'artificial' persons: Kelsen (1992: 48-9, 2005: 174-91).

²⁷⁷ Kelsen (1992: 7, c.f. 2005).

constructive interpretation’,²⁷⁸ that is, it ‘requires our judges [...] to treat our present system of public standards as expressing and respecting a coherent set of principles, and, to that end, to interpret these standards to find implicit standards between and beneath the explicit ones’.²⁷⁹ But according to legalism there is no reason to treat our present system of public standards as expressing and respecting any coherent set of underlying principles. So this attractive adjudicative principle cannot get off the ground.²⁸⁰ Because of this, Finnis critiques legalistic approaches on the basis that such approaches offer ‘no guidance’ for ‘judges confronted by the practical problems of identifying law and rights after revolution’.²⁸¹ Whilst legalists might be content to bite this bullet, it is worth noting that many of law’s subjects want to see some kind of principled consistency in legislation and adjudication. Indeed, when the European Parliament tabled the legislative proposal to create a new legal status as ‘electronic persons’ for certain autonomous robots, it was a central point of contention in the ensuing public debate that ‘[f]rom an ethical [...] perspective, creating a legal personality for a robot is inappropriate’.²⁸² This leads nicely into the second, more principled problem: as Finnis argues, legalistic approaches seem to mistake ‘law’s point’, namely that ascriptions of legal personhood should translate ‘disciplined and critical concern with the interests and well-being of real people’.²⁸³ How might legal personhood ascriptions reflect this kind of concern?

The school of thought concerning legal personhood which I call ‘personhood moralism’, or ‘moralism’ for short, centres on the idea that, as a general rule, legal personhood

²⁷⁸ Dworkin (1986: vii).

²⁷⁹ *Ibid*: 217. Dworkin (1986: 217) calls this the ‘adjudicative principle of integrity’.

²⁸⁰ This is one way in which Dworkin’s adjudicative principle of integrity depends upon his ‘legislative principle of integrity’, which ‘asks those who create law by legislation to keep that law coherent in principle’ (Dworkin 1986: 167). If the legislative principle is false, as it is according to legalism, it is unclear why judges should interpret law as morally coherent.

²⁸¹ Finnis (2011b: 26). In a similar vein, legalism quite closely resembles the ‘plain fact’ view of law which Dworkin (1986: ch.1) critiques because it cannot explain or support adjudication in ‘hard cases’ in law.

²⁸² Open Letter to the EU Commission (2018).

²⁸³ Finnis (2011b: 26). This echoes Noonan’s (1976) classic case that judges who reasoned legalistically, viewing legal persons as ‘masks’ rather than genuine ‘persons’, produced and reproduced the injustice of legal slavery for decades.

depends upon moral personhood.²⁸⁴ On the standard view, moralists assume that ascriptions of moral personhood—and, by extension, of *legal* personhood—depend upon an entity’s intrinsic properties, such as dignity or sentience.²⁸⁵ For instance, Finnis proposes a moralist conception of legal personhood according to which ‘all human beings [...] have the ‘right to recognition everywhere as a person before the law’ [UDHR, Art. 6]’²⁸⁶ due to the intrinsic dignity of every human being.²⁸⁷ As we saw above, moralism appears to be the working assumption in the SARs literature so far and the account of legal personhood which I defend in this section is a version of moralism.

But the standard assumption that moral personhood—and so legal personhood—is grounded in some intrinsic property like dignity or sentience does not fit with an important extensional fact about ascriptions of legal personhood, namely that courts across the Western world grant corporations legal personhood and have done so since the nineteenth century.²⁸⁸ As a result, although a corporation lacks dignity or sentience, it can nonetheless enter into contracts, sue and be sued, own property and so on. Moralists tend to view the legal personhood of corporations as ‘simply another mode by which *individuals* or *natural persons* can enjoy their property and engage in business’.²⁸⁹ Indeed, Finnis endorses this kind of position, arguing that, since ‘a corporation’s rights [...] are analysable without remainder into individuals’ rights’,²⁹⁰ we should understand the legal personhood of corporations as justified not by the

²⁸⁴ For some apparent exceptions: *supra* n.89.

²⁸⁵ For detailed discussion: Naffine (2009: 20). I bracket *rationalism*—viz., the view that legal personhood is grounded in an entity’s rationality—since this ostensibly under-extends, ruling out those natural persons who do not meet some standard of rationality, including the severely mentally ill, infants, those suffering neurological degeneration or comatose adults.

²⁸⁶ Finnis ([2000]2013: 23).

²⁸⁷ Although Naffine calls this kind of position ‘religionism’, the Universal Declaration of Human Rights makes clear that considerations of intrinsic dignity need not be grounded in any world religion.

²⁸⁸ See, e.g., Savigny (1840: §60, §85) (Germany); *Trustees of Dartmouth College v Woodward* [1819] 17 U.S. 518; *Santa Clara Co v. Southern Pacific Railroad* [1886] 118 U.S. 394 (U.S.A); Blackstone (1876: 455, 186-7) (England). For an overview of the historical corporate personality debate: Conard (1976: 417-45). C.f. Freund ([1897]2000); Dewey (1926); Machen (1911); Hallis (1930). Olivecrona (1928) provides a helpful summary of ‘classical’ Continental theories of the nineteenth century.

²⁸⁹ Hohfeld (1909: 197).

²⁹⁰ Finnis (2011b: 30).

corporation's intrinsic properties *per se* but rather, ultimately, by appeal to the dignity of individual human persons. But the problem remains: corporations, on the standard moralist view, are *legal* but not *moral* persons, which is in tension with the core claim of moralism.²⁹¹ Such considerations are not decisive, but they suggest that dismissing the legal personhood of SARs based upon their lack of certain intrinsic features is undermotivated. As the case of corporations shows, legal personhood need not depend upon those intrinsic properties. On what basis, then, should we assess the proposal to ascribe legal personhood to SARs?

4.1.2. Legislative Integrity

In *Law's Empire*, Dworkin proposes what he calls the 'legislative principle of integrity', which 'asks those who create law by legislation to keep that law coherent in principle',²⁹² that is, 'to try to make the total set of laws morally coherent'.²⁹³ On this view, there are two dimensions on which we should assess the legitimacy of any new legal proposition: firstly, does the proposition *fit*—i.e., is it *consistent*—with the norms conveyed by legal institutions? Secondly, is the proposition *justified* by (a) sound moral principle(s)? In light of the legislative principle of integrity, then, I want to reflect, in the context of SARs, on the European Parliament's proposal that at least some autonomous robots should be granted legal personhood. Integrity calls us to evaluate this proposal along two dimensions:

Q1. Would the ascription of legal personhood to SARs fit with the norms conveyed by legal institutions?

Q2. Would the ascription of legal personhood to SARs be justified by a sound moral principle?

²⁹¹ Finnis might respond that corporations are not central cases of legal personhood but merely peripheral ones. But, as I show, the legal personhood of corporations seems to be a basic extensional fact about personhood in contemporary legal systems, one not to be confined to the conceptual peripheries.

²⁹² Dworkin (1986: 167).

²⁹³ *Ibid*: 176.

To begin with, then, we need some account of the norms conveyed by legal institutions concerning ascriptions of legal personhood.

Because my thesis is primarily philosophical rather than legal, I will not offer a complete answer to Q1. Instead, I restrict myself to the stipulation of just two familiar extensional facts concerning ascriptions of legal personhood which any account of the legal personhood of SARs must fit and justify. I focus on extensional facts about legal personhood ascriptions in contemporary Western legal systems, including common law (the UK and US) and civil law (Germany, France, Italy, Spain, Austria) jurisdictions. This makes sense for reasons of space and the contingent truth that, as we have seen, contemporary political discussions of AI legal personhood concern primarily these kinds of systems. Two extensional facts about legal personhood in these legal systems are as follows:²⁹⁴

F1. All human beings who have been born, are currently alive and are sentient are legal persons.

This is because: only human beings are so-called ‘natural persons’;²⁹⁵ being born alive is generally sufficient if not necessary to warrant the ascription of legal personhood;²⁹⁶ dead people are not legal persons;²⁹⁷ it is uncontroversial that born, living, sentient humans are legal persons.²⁹⁸

F2. In addition to natural persons, there are so-called ‘artificial persons’, such as corporations.

²⁹⁴ For elaboration of these and other extensional facts in contemporary Western legal systems: Kurki (2019: *Introduction*).

²⁹⁵ See: Berg (2007: 373); c.f. German Civil Code (§1); Austrian Civil Code (§16).

²⁹⁶ Being born is explicitly necessary for natural personhood in French, German and Italian law. See, e.g., German Civil Code §1; Italian Civil Code Pt.1:§1. In Spanish law, the neonate must also live for 24 hours outside of the womb (de las Heras Ballell 2010). An American landmark case, defining foetuses as non-persons under the 14th Amendment of the Constitution was *Roe v. Wade* [1972] 410 U.S. 113.

²⁹⁷ Although Kramer (2001) suggests that dead people might possess some limited rights. C.f. Stemplowska (2020).

²⁹⁸ There is controversy regarding the legal personhood of anencephalic infants since they lack sentience. Anencephalic infants are legal persons in the US (Berg 2007: 377-379) and Germany (Neuner 2013: 651).

In general, there appear to be three principal ways in which the legal personhood of corporations has historically been understood:²⁹⁹

- (i) *Concession (or fiction) theories*: corporations exist only as creations of the state or the sovereign.
- (ii) *Participant theories*: corporations have legal rights and/or legal duties which encode the legal rights and/or duties of the individual legal persons who compose the corporation.³⁰⁰
- (iii) *Real entity theories*: a corporation is a real entity over-and-above its members and has legal rights and/or duties in a given domain in virtue of the role that it plays in that domain.

The classic judgment cited as laying out the concession theory of the corporation is *Trustees of Dartmouth College v Woodward*, which states that corporations exist only in the ‘contemplation of law’.³⁰¹ More generally, as Kurki shows,³⁰² concession theories were popular until the middle of the nineteenth century.³⁰³ But concession theories were largely supplanted by participant theories in the late nineteenth century. For example, in *Santa Clara Co v. Southern Pacific Railroad*, the Court is clearly ‘looking through the corporation to its shareholders and their constitutional rights’.³⁰⁴ The last century has seen a decisive and significant deepening of the legal status of corporations, endowing corporations with a status that seems to me best described by real entity theories. This interpretation of legal personhood has, for example, been the dominant one in the US Supreme Court in recent years, which ‘attribute[s] constitutional

²⁹⁹ Orts (2013).

³⁰⁰ C.f. Jones (1999) on the ‘collective conception’ of groups.

³⁰¹ [1819] 17 U.S. (4 Wheat.) 518: 636.

³⁰² Kurki (2019: ch.1).

³⁰³ Blackstone (1876: 186-7); Savigny (1840: §85). Concession theories have an even longer pedigree: in 1246, Pope Innocent IV declared that corporations were *persona ficta* (List and Pettit 2011: 170).

³⁰⁴ Finnis (2013: 46, n.4). See also Kelsen’s theory of corporate legal personhood, on which the legal personhood of corporations is no more a legal fiction than is any other kind of legal personhood and is entirely explainable in terms of its role in regulating human conduct in law.

rights to corporations precisely as real entities in themselves’.³⁰⁵ Indeed, the Supreme Court has ruled that business corporations have the same free speech rights as natural persons in election campaigns;³⁰⁶ religious freedom under a federal law not to provide birth control coverage in employee health plans³⁰⁷ and, arguably, certain religious freedom rights under the First Amendment.³⁰⁸ As a result, I understand the legal personhood of corporations signalled in F2 according to a real entity theoretic approach.³⁰⁹

Based upon my arguments in 4.1.1, it appears that legalism *fits* but does not *justify* F1 and F2. Meanwhile, the standard interpretation of moralism can *justify* F1 and F2, but does not *fit* them both. With these considerations in mind, how should we interpret F1 and F2?

4.1.3. Personhood Functionalism

One way to amend the moralist account of legal personhood in order to better accommodate F2 would be to ground moral personhood in something other than dignity or sentience. This is what List and Pettit do in their functionalist account of moral personhood as *moral agency*: ‘a [moral] person is an agent who can perform effectively in the space of moral obligations’.³¹⁰ As I unpack this definition, I explain why certain groups, which I call corporations, satisfy it.

Firstly, an agent is defined in the following way. For some entity X and some context C, X is an agent in C iff:³¹¹

A1.X has representational states that depict how things are in the environment in C;

³⁰⁵ *Ibid*; c.f. Horwitz (1985: 178, 192, 189).

³⁰⁶ *Citizens United v. Federal Election Commission* [2010] 558 U.S. 310.

³⁰⁷ *Burwell v. Hobby Lobby Stores, Inc.* [2014] 573 U.S. 682.

³⁰⁸ *Masterpiece Cakeshop, Ltd. V. Colorado Civil Rights Commission* [2018] 584 U.S. 617.

³⁰⁹ Considerations advanced in the next subsections in favour of a functionalist conception of legal personhood offer further support for the conceptualisation of corporations as real entities over and above their individual members.

³¹⁰ List and Pettit (2011: 173).

³¹¹ *Ibid*: 19-20.

A2.X has motivational states that specify how X requires things to be in that environment;

and

A3.X has the capacity to process its representational and motivational states, leading it to intervene suitably in the environment in C whenever that environment fails to match a motivating specification.

List and Pettit argue that different organisational structures in groups of people mean that certain groups—which I denote by the term ‘corporations’—are agents in certain contexts. First, corporations have mechanisms for forming representational states. For example, some corporation Y’s public relations team may collect polling data which informs the judgment ascribed to the corporation as a whole that its latest publicity campaign is transphobic. Second, corporations have motivational states, that is, preferences about how the group requires the state of affairs to be. Y’s constitution, for instance, might require that all publicity campaigns comply with an ethos of inclusivity. Third, corporations have decision-making mechanisms which endow the group with the capacity to intervene suitably in a given situation on the basis of its judgments and preferences about that situation: Y might discontinue the publicity campaign in question on the basis that its executive committee has voted against it.

Secondly, to perform effectively in the space of moral obligations, an agent must satisfy a further three conditions, which are necessary and sufficient for the agent to ‘contract[...] obligations to others and [...] derive[...] entitlements from the reciprocal obligations of others’.³¹² For some agent X in C, X is a moral person in C iff:³¹³

A4.X faces a normatively significant choice in C;

A5.X has the understanding and access to evidence required for making judgments about the options in C;

³¹² List and Pettit (2011: 173).

³¹³ *Ibid*: 155, ch.8.

A6.X has the control required to make decisions in C.

For example, Y's accounting team might inform Y's judgment that the transphobic publicity campaign is exceptionally profitable. Suppose that Y's constitution contains a clause allowing that, when they are exceptionally profitable, the discontinuation of publicity campaigns non-compliant with an ethos of inclusivity is a matter for the discretion of Y's executive board. Y therefore faces a normatively significant choice between retaining a transphobic, profitable campaign or discontinuing the campaign in favour of an alternative non-transphobic one. Y's policy analysis team informs Y's judgment that (1) the current campaign is exceptionally profitable *because* it is transphobic and (2) it is therefore overwhelmingly likely that an alternative campaign would be less profitable but (3) there is good evidence to suggest that corporations which consistently display inclusive values in their public relations out-compete others over ten-year time frames. Y's executive board therefore votes on the options and, on this basis, Y decides to continue the campaign. This illustrates one way in which a group agent performs effectively in the space of *moral* obligations.

The functionalist account of legal personhood which I propose grounds the legal personhood of some entity in its moral agency, understood according to the functionalist approach glossed above. In other words, a legal person is a moral agent which is able to perform effectively in the space of *legal* obligations. To complete the illustration of my proposal, a claimant, Z, could meaningfully bring a legal case against Y, claiming that the campaign violates their legally protected right in that jurisdiction to be protected from discrimination on the basis of their gender identity. This right correlates to a duty borne by Y not to engage in certain forms of speech or expression. Y's defence might hinge upon its legally protected right, as a legal person, to freedom of speech or expression. The court—another kind of corporation—might then need to decide whether to grant Z's right or Y's right precedence in this context.

Such considerations are of course not decisive, but I believe they give us sufficient reason to grant that the functionalist definition of legal personhood fits F2.

One might worry, however, that the functionalist interpretation does not fit F1 since not all natural persons perform effectively in the space of legal obligations. Think, in particular, of those who only bear what is called ‘passive’ legal personhood, for instance infants.³¹⁴ To resolve this concern, I will gloss what ‘passive’ and ‘active’ legal personhood standardly include. The Hohfeldian incidents paradigmatically associated with passive legal personhood include certain fundamental claim-rights (protections of life, liberty and bodily integrity), the capacity to own property, victim status in criminal law and legal standing to enforce one’s claim-rights in court through a representative. The Hohfeldian incidents paradigmatically associated with ‘active’ legal personhood—a legal status which typically accrues to human adults of sound mind—include criminal-law and tort-law responsibility and the capacity to enter into contracts.³¹⁵ But it appears to me that the incidents associated with passive legal personhood are sufficient for an entity to perform effectively in the space of legal obligations (albeit primarily as rights-holders rather than duty-bearers). So the functionalist interpretation of legal personhood seems to fit F1, too.

Another objection, however, is that functionalism achieves fit with F1 and F2 only at the cost of emptying the concept of legal personhood of all its normative content. Indeed, one might ask in what sense the functionalist account of legal personhood actually *justifies* any ascriptions of legal personhood since it does not depend upon any intrinsic, *morally relevant* features of an entity, such as dignity or sentience. My response has two parts. Firstly, a note of concession. To some extent, it is true that the functionalist account of legal personhood does not depend upon the entity to which it is ascribed having any *intrinsic* moral status, that is,

³¹⁴ Kurki (2019: 94-96); MacCormick ([1988]2007: 86).

³¹⁵ Kurki (2019: 94-96); MacCormick ([1988]2007: 86, 93).

moral status which an entity has independently of its relationships with other entities. To illustrate, a corporation has no intrinsic moral status but a human baby does: destroying the former, for example, seems to be wrong only in virtue of the effects which that destruction has on entities like natural persons who may lose their jobs and suffer; destroying the latter, by contrast, is wrong in itself. But secondly—and this is crucial—my account *does* depend upon the entity to which it is ascribed having at least *extrinsic* moral status, that is, moral status which an entity has in virtue of its relationships with other entities. In the former example, the corporation seems to have extrinsic moral status in virtue of its relationships with various intrinsically morally valuable humans. The functionalist conception of legal personhood has not emptied legal personhood of its normative content; on the contrary, functionalism promises an account of legal personhood which is sensitive to the normative significance of our relationships with non-human entities. In order to further develop this latter point, I wish to clarify the normative content of the functionalist conception of legal personhood through applying the theory developed in this section to the case of SARs.

4.2. Oh Brave New World: Towards a Legal Framework for SARs

In the previous section, I articulated and defended a functionalist conception of legal personhood, one which best satisfies the dimensions of *fit* and *soundness* on which the ideal of integrity evaluates the development of law. In this section, I show that SARs should be legal persons according to this functionalist conception (4.2.1). In 4.2.2, I distinguish between non-derivative and derivative legal rights—that is, respectively, legal rights which are grounded in the rights-holders own interests and those which are grounded in the interests of others—and argue for the ascription of derivative rights against robotic sexual violence to SARs. 4.2.3 defends my response to the robotic sexual violence problem against certain objections.

4.2.1. The Legal Personhood of SARs

To begin with, I argue that SARs are *agents*, that is, they satisfy (A1)-(A3).³¹⁶ There appear to be two principal ways to justify this claim. Firstly, by drawing an analogy with the case of corporations. To see how this works, consider Dennett's 'intentional stance' test for agency,³¹⁷ according to which '[a]nything that is usefully and voluminously predictable from the intentional stance is, by definition, an intentional system'.³¹⁸ Theorists of group agency have argued that passing the intentional stance test is good evidence that a system satisfies (A1)-(A3).³¹⁹ For instance, the kinds of entities which we count as artificial persons in law (e.g. firms, corporations and other organised collectives) are 'usefully and voluminously predictable from the intentional stance'. In addition to this, List shows that any AI system which passes a suitably specified version of Turing's 'imitation game' test will satisfy Dennett's intentional stance test.³²⁰ Specifically, Turing proposed that when assessing whether a system is 'intelligent', we should replace the question 'Can machines think?' with the more tractable question 'Are there imaginable digital computers which would do well in the imitation game?', where this means displaying behaviour that is indistinguishable, in an observer's eyes, from that of an intelligent agent.³²¹ Different specifications of the imitation game correspond to different versions of this test. In Turing's own version, the system must imitate the responses of a human interlocutor in written communication via an instant-messaging system. ChatGPT seems to satisfy this version of the test. Because of the more intimate context of their use, the appropriate version of the Turing test for SARs would require more sophisticated capabilities.

³¹⁶ Indeed, the illustrative example of agency in List and Pettit (2011) is a much simpler robot than a SAR. List and Pettit ask us to consider a 'small robotic device' which moves around on a table-top covered in cylinders, some upright, some on their sides. Suppose that the robot 'scans the area with its bug-like eyes, then moves about and, using its prosthetic limbs, puts any cylinders lying on their side into an upright position' (2011: 19). List and Pettit claim that the robot satisfies conditions (A1)-(A3) and is, therefore, an intentional agent in this context.

³¹⁷ Dennett (1987, 2009).

³¹⁸ Dennett (2009: 339).

³¹⁹ For arguments to this effect, see: Tollefsen (2002, 2015); List and Pettit (2011).

³²⁰ List (2021).

³²¹ Turing (1950: 442).

But, as I explained in chapter one, whilst autonomous robots which adequately mimic the behaviour of a real human sexual partner do not yet exist, the requisite technology and corporate interest in their development seem to. As a result, I submit that SARs would satisfy a suitably specified imitation game which, as List points out, offers good reason to infer that they would pass Dennett's intentional stance test and, as a result, satisfy (A1)-(A3).

Secondly, List argues that group agents like corporations are not just *analogous* to but rather *special cases* of AI systems, where the 'hardware' supporting the AI is social rather than electronic.³²² This draws on Parikh's concept of 'social software', which captures the idea that certain social phenomena are appropriately conceptualised as computational processes implemented in social systems.³²³ Parikh uses this term to refer to a broad range of social phenomena, including institutional decision-making and voting mechanisms, but he does not focus on group agency. List extends the notion of social software, arguing that 'group agents are socially implemented AI systems'.³²⁴ Indeed, as functionalist philosophers of mind have long noted, agency is a multiply realisable phenomenon. In this vein, List contends that agency admits at least three different kinds of 'hardware systems':

- Biological systems (e.g., in human and non-human animals which display agency);
- Electronic systems (e.g., in AI systems such as autonomous robots);
- Social systems (e.g., in the case of group agents).

List notes that the boundaries between these different kinds of hardware may become increasingly blurred with the arrival of controversial new technologies such as biological robotics, human enhancement and technologically augmented social systems. But such

³²² List (2021).

³²³ Parikh (2002).

³²⁴ List (2021: 1221).

considerations need not detain us here since I believe that my arguments so far suffice to show that SARs, like group agents, would satisfy (A1)-(A3).

In addition to this, SARs appear to be *moral* agents, that is, agents which perform effectively in the space of moral obligations. To illustrate, consider a SAR in a sexual context. Because I have argued that we may take the intentional stance towards the robot, I will use phrases such as ‘the SAR judges/intends/wants/acts’. Suppose that the SAR, in this context, judges that its human user is aroused, aggressive and attempting to initiate a sexual encounter on the basis of the user’s body language, speech and facial expressions. Since it is autonomous, the SAR faces the choice as to how to respond: should it consent enthusiastically, acquiesce, say no or shut down? Any one of these choices, as I showed in the last chapter, is normatively (if not yet legally) significant. As a result, the SAR satisfies (A4). The SAR understands that its user’s behaviour is aggressive and has sufficient evidence from past sexual encounters to judge that saying no or shutting down increases the risk of damage to its systems, whilst signalling acquiescence or consent encourages the same kind of future behaviour in the user. As such, the SAR seems to satisfy (A5). Based upon its internal evaluation of what would minimise harm to its systems in the long term, the SAR determines its response, perhaps saying ‘no’ to its user’s sexual advances. This seems to satisfy (A6). I submit that SARs, then, seem to be moral agents according to the functionalist conception of moral agency. According to the functionalist interpretation of legal personhood, which I defended in the last section, SARs may also, therefore, be meaningfully ascribed legal personhood. In other words, SARs would be appropriate holders of legal rights or bearers of legal duties. In the next subsection, I turn to the consideration of which kinds of legal rights SARs may appropriately bear; I defer consideration of the possibility that SARs might also bear certain kinds of legal responsibilities until the next chapter.

4.2.2. Robot Rights

As I explained in the last section, according to the functionalist conception which I have defended, the ascription of legal personhood to some entity does not depend upon the intrinsic moral status or value of that entity. Entities, such as corporations, which lack intrinsic moral status may therefore be legal persons in virtue only of their extrinsic moral status. The distinction between intrinsic and extrinsic moral status sustains a second distinction between non-derivative and derivative legal rights. Non-derivative legal rights are those which a legal person has in virtue of her intrinsic moral status. For example, the kinds of fundamental protections which all natural persons have for their lives, liberty and property are non-derivative rights. Non-derivative rights are typically justified by appeal to something like the moral value of certain interests of the rights-holder: X's non-derivative rights protect those interests of X which are important enough to generate legal duties on others to ensure their protection. Derivative legal rights, on the other hand, are those which a legal person has in virtue of the kinds of values or goods that they promote. For example, the property rights of corporations are typically justified by appeal to the desideratum of an efficient system of economic exchange.³²⁵ In this subsection, I argue for the ascription of derivative legal rights to SARs against subjection to robotic sexual violence.

To begin with, I assume that SARs do not possess the kinds of intrinsic features which are generally believed to underpin intrinsic moral status, features such as dignity or sentience.³²⁶ This is because, as I explained in chapter one, SARs involve only narrow AI and not that general AI which might be, in some sense, 'conscious'.³²⁷ As a result, any legal rights which we ascribe to SARs will be *derivative*, that is, justified by appeal to the values and goods

³²⁵ Turner (2019).

³²⁶ In the next chapter, I relax this assumption to briefly explore the kinds of normative issues which would be raised by the more futuristic possibility of artificial general intelligence.

³²⁷ Danaher (2017a); Eskens (2017).

which those rights promote. The central valuable good on which I wish to focus is that which I showed in the last chapter was a central matter of concern for the basic structure of society, namely reciprocity. I argued that robotic sexual violence undermines reciprocity because it undermines measures designed to ensure adequate all-purpose means for all citizens to make effective use of their freedoms. Because of this, I argued that reciprocity depends upon the elimination of robotic sexual violence. The legal proscription of robotic sexual violence, as a result, was shown to be part of reasonable measures for the cultivation of political virtue. One way to do this, I propose, is to create a legal right of SARs to be protected from subjection to robotic sexual violence, a derivative right justified by appeal to its efficacy in protecting the valuable good of reciprocity. The question, then, becomes whether the ascription of such a legal right to SARs would, indeed, be an efficacious way to achieve this valuable social end.

There are several ways to justify the view that the ascription of this legal right to SARs would support reciprocity. In general, one might appeal to the positivist assertion that law, insofar as it claims to be law, asserts authority to give its subjects content-independent reasons for action.³²⁸ As a result, there is something intrinsically action-guiding—albeit not necessarily *legitimately* so—about the ascription of a legal right to SARs proscribing some form of conduct against it. Burge-Hendrix draws attention to a certain ‘educative’ function of the law implicit in this positivist thesis, according to which part of the law’s function is to ‘make its subjects aware of the set of norms that structure[...] social and political life, and to persuade them to follow those norms’.³²⁹ Indeed, in his discussion of how social norms may solve collective action problems, Sunstein contends that the law may be used ‘as an effort to produce adequate social norms’, even when unaccompanied by much in the way of enforcement.³³⁰ He takes as

³²⁸ Raz (1994).

³²⁹ Burge-Hendrix (2007: 243).

³³⁰ Sunstein (1996: 2031).

an example laws which forbid littering and require that people clean up after their dogs. Although such laws are rarely enforced through the criminal law,

‘they have an important effect in signalling appropriate behavior and in inculcating the expectation of social opprobrium and, hence, shame in those who deviate from the announced norm’.³³¹

As a result, with or without enforcement, laws may ‘help to reconstruct norms and the social meaning of action’.³³² In a similar way, although there might be practical or principled difficulties with enforcing the legal proscription on robotic sexual violence in criminal law, as I discussed in the last chapter, ascribing a legal right to SARs to be protected from robotic sexual violence may help to reconstruct sexual norms and the social meaning of certain violent sexual behaviours in ways which would protect the basic rights of women and girls. This kind of thinking, moreover, underlies Green’s case that law may be used to shape social norms and, in particular, to help foster better sexual norms. But he writes:

‘thinking that *only* rape law stands a chance of shaping attitudes to rape is no better than jurisprudential homeopathy. In fact, it may be that the most influential laws in shaping our social morality here will not be the laws of rape at all [...] We should not assume that only like cures like: a wide variety of legislative initiatives can shore up the background conditions we are concerned about’.³³³

In other words, we are called to be not only open-minded in our consideration of which legal measures might help to foster positive social change; we are also asked to be thoroughgoing in our commitment to ensuring that legislative initiatives in all areas of life, including the most intimate ones, support the development of social norms conducive to reciprocity. My proposal

³³¹ *Ibid*: 2032.

³³² *Ibid*.

³³³ Green (2013: 492).

is that the ascription of legal rights to SARs proscribing robotic sexual violence is supportive of this valuable aim.

4.2.3. Objections and Replies

One objection to my argument for the ascription to SARs of a legal right against robotic sexual violence appeals to the standard, interest theory of legal rights. The most oft-cited version of this theory is Raz's:

“X has a right’ if and only if X can have rights, and, other things being equal, an aspect of X’s wellbeing (his interest) is a sufficient reason for holding some other person(s) to be under a duty’.³³⁴

Raz continues: ‘[a]n individual is capable of having rights if and only if either his wellbeing is of ultimate value or he is an ‘artificial person’ (e.g. a corporation).’³³⁵ I have argued that autonomous robots like SARs should be artificial persons. But this leaves two questions. First: in what sense could a SAR be said to have ‘wellbeing’ or ‘interests’? Of course, if *consciousness* or *sentience* is required for an entity to have wellbeing or interests, then the ascription of legal rights to SARs, according to the Razian interest theory, is dead in the water. But as I explain below, Raz does not believe that wellbeing or interests—at least for *artificial* persons—depend upon these kinds of intrinsic features. Second, then, even if the first question may be answered, how could a *robot*’s interest be ‘a sufficient reason’ for holding *humans* to be under a corresponding duty?

To answer the first question, I draw again on the analogy between SARs and artificial persons such as corporations. The sense in which *corporations* have interests on the Razian account is intrinsically tied up with their legal rights and duties:

³³⁴ Raz (1986: 166; 1994: 266–269).

³³⁵ Raz (1986: 166).

‘Corporations also have interests determined by their purposes, powers, and duties. It is true that protecting these interests is not intrinsically valuable. Nevertheless, corporations [...] have rights in the same sense as other individuals. They have rights if and only if their interests are sufficient to justify holding others to be subject to duties’.³³⁶

One interpretation of what Raz says here is to define corporate rights in terms of corporate ‘purposes, powers, and duties’. But this would be odd because it seems to misconstrue the justificatory story of corporate rights. As Pallikkathayil explains: ‘artificial persons [...] are created as packages of powers, rights, and duties, with each of these tailored to fit the others’ such that ‘[e]xplaining the rights in terms of the duties [...] miss[es] the way in which the rights and duties are justified altogether (or not at all)’.³³⁷ This kind of interpretation of Raz captures the interdependence of legal rights and duties according to the functionalist conception of legal personhood: that corporations, like SARs, perform effectively in the space of legal obligations—that is, they are legal persons—is inseparable from the ascription to them of certain legal rights and duties. But the *interests* of artificial persons like corporations and, I suggest, SARs in achieving certain ends do not seem, on the Razian account, to be explicable independently of their legal rights and duties. This prompts the next question, which concerns the value of interests like these.

To answer the second question, I will briefly explain how certain kinds of *instrumentally* valuable interests which accrue to legal persons on Raz’s account may be sufficient to ground a duty. As Raz explains:

³³⁶ Raz (1995: 259).

³³⁷ Pallikkathayil (2016: 149). I return in the next chapter to the question of what legal duties SARs might bear.

‘[r]ights protect the interests of right-holders. But these interests need not be intrinsically valuable. The reason for protecting them may be that by doing so one does protect the interests of others’.³³⁸

He elaborates on this idea of instrumentally valuable interests as follows:

‘The rights of journalists [...] to protect their sources are normally justified by the interests of journalists in being able to collect information. But that interest is deemed to be worth protecting because it serves the public. That is, the journalists’ interest is valued because of its usefulness to members of the public at large. The rights of priests, doctors and lawyers to preserve the confidentiality of their professional contacts are likewise justified ultimately by their value to members of the community at large’.³³⁹

There are well-known concerns about the compatibility between this account of instrumentally valuable rights with the rest of Raz’s interest theory.³⁴⁰ Of course, if Kamm is right in her objections, the internal inconsistency of the interest theory would leave the objection against my account considered here undermotivated. Pallikkathayil, at any rate, argues that Kamm’s critique fails when we attend appropriately to the distinction between intrinsically and instrumentally valuable rights. Indeed, Pallikkathayil proposes that we interpret Raz’s interest theory thus: “X has a right” if and only if X can have rights, and, other things being equal, the intrinsic or instrumental value of an interest of X’s is a sufficient reason for holding some other person(s) to be under a duty’.³⁴¹ On this understanding of the interest theory, my contention in Part One of this dissertation may be understood as follows: the instrumental value of the right of SARs against robotic sexual violence is a sufficient reason to hold the users of SARs to be under a duty to refrain from robotic sexual violence. Based on my arguments in the previous

³³⁸ Raz (1995: 259).

³³⁹ *Ibid.*

³⁴⁰ Kamm (2002).

³⁴¹ Pallikkathayil (2016: 150).

chapter, which established that robotic sexual violence is a matter of concern for the basic structure of society, I believe that this contention is well-grounded.

A follow-up objection to my arguments in this section asks why, if I can convey my whole thesis in terms of derivative legal rights, I have considered it valuable and instructive to discuss legal personhood for SARs at all. Of course, the orthodox view is that all and only legal persons either hold legal rights and bear legal duties, or at least have the legal capacity to hold rights and bear duties.³⁴² As a result, the ascription of legal personhood to SARs is necessary for the ascription to them of any legal rights. But if legal personhood is simply a shorthand for legal rights and duties and if, as I have argued, the legal rights and duties of SARs are only instrumentally valuable, did we really need all this talk of legal personhood in this chapter?³⁴³ There are two reasons why we did. Firstly, the ascription of legal personhood to SARs is not just a necessary condition for the ascription of legal rights and/or duties. Rather, there is a flexibility and an open-endedness to ascriptions of legal personhood which leaves open the possibility of ascribing to the entity new legal rights and responsibilities as the function of that entity in society changes with time. This sometimes controversial possibility of broadening and deepening the nature and scope of a legal person's rights is adverted to by the recent series of US Supreme Court rulings concerning the legal rights of corporations which I alluded to in 4.1.2. I do not purport in Part One of this dissertation to offer an exhaustive list of the legal rights and responsibilities of SARs and have focused only on the robotic sexual violence context. As a result, the flexibility and open-endedness of the ascription of legal personhood is my invitation to future scholarship to take the project further.

³⁴² *Supra* n.14.

³⁴³ DeGrazia (2008: 184) believes that 'talk of moral [personhood] is redundant and can always in principle be replaced by other language'. Stone (1972[2010]: 22) considers a similar objection to his proposal that natural objects, such as trees, should have the kind of legal status which he calls 'standing'.

Secondly, and more subtly, as I explained above, there is an important expressive function of certain legal statements which serves to shape human attitudes, beliefs and behaviours in valuable ways. In the context of the environmental movement, Stone has adverted to the valuable ‘psychic and socio-psychic aspects’ of ascribing legal personhood to certain kinds of natural objects, reflecting and sustaining a social norm which values the environment in its own right.³⁴⁴ Whilst ascriptions of legal personhood to new entities are often justified by appeal to the interests of existing legal persons,³⁴⁵ the expressive function of the law means that the legal personhood of natural objects encourages us to interact with the natural world with care and responsibility. In a similar way, the ascription of legal personhood to autonomous robots such as SARs may encourage norms of sexual engagement which embody care and responsibility, something which may turn out to be valuable not just to women and girls but also, incidentally, as a more intrinsically valuable way of being.

A third and final kind of concern which one might have at this point in my argument is as follows. I have articulated and defended a novel, public reasons argument for the legal proscription of robotic sexual violence, a legal proscription which may fruitfully be implemented through the legal personification of SARs. Although robotic sexual violence represents a more pressing concern given the risks associated with autonomous AI, my argument contained as a corollary a wholly new public reasons argument for the legal proscription of pornographic sexual violence.³⁴⁶ This prompts the following questions. Firstly, would my arguments justify the legal proscription of AI-generated pornographic sexual violence, since this would implement the same kinds of machine learning technologies which

³⁴⁴ Stone (1972[2010]: 23-31).

³⁴⁵ *Ibid*: 24.

³⁴⁶ Watson (2007) is to my knowledge the only other author to entertain the idea of a public reasons justification for the legal regulation of pornography. She offers a Rawlsian reimagining of MacKinnon’s arguments for its legal proscription. Significant differences between our approaches include that Watson neither analyses reciprocity, nor relates public reason to epistemic injustice.

make SARs particularly pernicious from the perspective of public reason? Secondly, if so, would my arguments similarly justify the legal personification of AI-generated pornographic sexual violence? In brief: my answer to the first question is *yes* and my answer to the second is *no*.

To elaborate briefly on my answer to the first question. In the US context, there has been growing concern about the appropriate regulation of AI-generated pornographic sexual violence, especially AI-generated child pornography.³⁴⁷ This type of pornographic material does not seem to cause legally recognised harms to any person given the Supreme Court's decision in *Ashcroft v. Free Speech Coalition*.³⁴⁸ One kind of response is to argue that there is a politically liberal justification for the legal proscription of AI-generated child pornography based upon direct or indirect harms caused to children.³⁴⁹ My arguments in Part One of this dissertation shine light on a new kind of politically liberal response, namely that AI-generated pornographic sexual violence, including AI-generated child pornography, is a matter of concern for the basic structure of society because it undermines reciprocity; additionally, arguments for its legal proscription can be presented independently of any comprehensive doctrine in terms of ideas which are implicit in the public political culture of a democratic society. A public reasons justification for the legal proscription of AI-generated pornographic sexual violence would seem to derive from the same considerations which I identified in developing the public reasons argument for the legal proscription of robotic sexual violence. So, I suggest that my arguments may contribute to the legal philosophical discourse concerning the regulation of AI-generated pornography.

³⁴⁷ See: Goldblatt (2012); Sieg (2005); Mains (2010). Recent statistics indicate growth in internet child pornography due to AI-generated content and problematic training data: Wong (2026).

³⁴⁸ 535 U.S. 234, 241 (2002). C.f. Rissman (2025).

³⁴⁹ Wilder (2024).

However, it does not follow that we should legally personify AI-generated pornography. This is because of certain important disanalogies between AI-generated pornography and SARs. I wish to briefly highlight three such disanalogies which mean that AI-generated pornographic sexual violence warrants a different regulatory approach. Firstly, AI-generated pornography is not a candidate for legal personhood because it does not satisfy the conditions of the functionalist account which I have defended. In brief, AI-generated pornography does not satisfy the conditions for agency, moral agency or effective performance in the space of legal obligations. Secondly, even if AI-generated pornography were a candidate for legal personhood, the expressive and educative benefits of legal personification which I adverted to above to in the context of SARs would not be engaged, or only to a lesser extent, for AI-generated pornography. In particular, because AI-generated pornography would not function as a substitute for a human sexual partner in anything like the same way as a SAR, the psychic and socio-psychic benefits adverted to above of legally personifying SARs do not apply to AI-generated pornography. Thirdly, the legal proscription of AI-generated pornographic sexual violence is enforceable by similar procedures to those applicable in the enforcement of criminal law prohibitions on ‘extreme pornography’ and child pornography.³⁵⁰ Given their function as something closer to a human partner, the same is not true for SARs. This is not, of course, a decisive argument against the legal personification of AI-generated pornography. However, it sketches some respects in which SARs generate distinctive challenges, which prompt a distinctive legal response.

4.3. Concluding Remarks

In this chapter, I have reflected on the proposal with which I opened Part One of this dissertation, namely the idea tabled by the European Parliament that at least some autonomous

³⁵⁰ For a helpful review of ongoing developments in UK law and policy in this vein: Pangotra (2025).

robots should be granted legal personhood. I have argued for a functionalist interpretation of personhood moralism (4.1), a position which supports the ascription of legal personhood to SARs, including derivative legal rights against subjection to robotic sexual violence (4.2). This completes my response to the robotic sexual violence problem.

5. Summing Up

*Ending is better than mending.*³⁵¹

5.1. Closing Considerations

I wish to close Part One of this dissertation by considering three questions which my arguments so far might evoke. The answers which I offer to these questions make no pretence to offer closure. But the thesis would seem wanting were I to offer no reflections on these matters.

Firstly, there is another kind of sexual violence which the emergence of SARs will make possible, one which reverses the roles played by human and SAR in the robotic sexual violence cases which I have focused on hitherto. For example, the development of ‘male’ SARs would make possible the following, disturbing case:

Reverse Robotic Rape: A tried-and-tested SAR, B, penetrates with its ‘penis’ the vagina anus or mouth of a human, A. A has not indicated consent. This is not a mechanical malfunction and no human intended or foresaw it; rather, it results from the SAR’s machine learning.³⁵²

It appears that, had a human committed the act, they would be charged with rape. The distinctive problem which is alleged to accrue to machine learning systems such as SARs is that there appears to be a morally problematic ‘responsibility gap’ when these systems cause wrongful harms.³⁵³ Some have argued that all those who contribute to the production of the autonomous system should bear strict civil or criminal liability for even unforeseeable damages which the system might cause.³⁵⁴ Another approach is to introduce faultless civil law

³⁵¹ Huxley (2002: 32).

³⁵² To illustrate: one deep-learning system was trained to reliably distinguish pictures of wolves from pictures of huskies; when it misclassified very clear images, it became apparent that the system had learnt to use the presence of snow in the picture as a determining factor (Ribeiro *et al.* 2016). Likewise, SARs might be overwhelmingly successful at recognizing signs of consent but for the wrong reasons, such that they go wrong in surprising cases.

³⁵³ For examination and critique of the very idea of a ‘gap’ in moral and/or legal responsibility: *infra* Part Two.

³⁵⁴ Burri (2017); Heyns (2013); Corn (2014); Crootof (2016). For one account of how autonomous robots could, at least in principle, be *punished* in criminal law: Abbott and Sarch (2019).

compensation schemes for damages caused by AI systems. Schellekens believes that this could replace traditional liability schemes for self-driving vehicles.³⁵⁵ Some have expressed concern that either of these schemes would have a ‘chilling effect’ on AI innovation.³⁵⁶ A third solution, as I explained at the start of Part One of this dissertation, is to introduce legal personhood for AI systems, à la *European Parliament resolution on Civil Law Rules on Robotics*, ‘so that at least the most sophisticated autonomous robots could [be legally] responsible for making good any damage they may cause’.³⁵⁷ My arguments in the previous chapter might seem to support this latter approach to the regulation of cases such as Reverse Robotic Rape. I defer further examination of this question of legal responsibility for autonomous AI until Part Two.

Secondly, then, you might be wondering how this kind of legal personhood for SARs would play out in practice. Turner glosses a variety of possibilities but the most promising one appears to be to house the legal personhood of AI systems like SARs within existing legal structures governing corporations.³⁵⁸ Bayern, for example, has argued that the US law of limited liability companies (LLCs) could be utilised to bestow legal personhood on any autonomous systems.³⁵⁹ In principle, he shows, it is possible to create an LLC where the operating agreement placed it under the control of an AI system and then have every other member of the LLC withdraw, leaving the system unsupervised by humans. In practice, there may be problems with housing AI legal personhood within existing legal frameworks given that this may not accord with legislative intention.³⁶⁰ But the attractions of the principle are manifold: as Turner explains, treating AI systems rather like LLCs would not only encourage

³⁵⁵ Schellekens (2018). Santoni de Sio and Mecacci (2021: 1078) criticise this approach. The context of self-driving vehicles generates legal controversy. Recently, when a self-driving car ran over a pedestrian, the onboard operator was considered negligent *and* liable for the death (BBC 2020). This approach strikes many as mistaken (Bellon 2018; Santoni de Sio and Mecacci 2021: 1078; Stilgoe 2020).

³⁵⁶ Trudy *et al.* (2021: 272).

³⁵⁷ Delvaux (2017: 18).

³⁵⁸ Turner (2019).

³⁵⁹ Bayern (2016).

³⁶⁰ Scherer (2017).

innovation—through acting as a valuable ‘firewall’ between existing legal persons and damages potentially caused by autonomous systems—but also allow AI systems to hold property, something which would resolve issues raised concerning the ownership of new intellectual property of which AI is the creator.³⁶¹

Finally, some might be curious as to how the development of general AI might alter the kind of legal framework which I have sketched in Part One of this dissertation. As I explained in chapter three, general AI would display humanlike intelligence and, as a result, would not be that motor for the replication and entrenchment of existing bias which narrow AI evinces. But equally, as I explained in chapter four, general AI would seem to possess at least some of the features commonly associated with *intrinsic* moral status, including sentience. As a result, general AI appears to be an appropriate candidate for the ascription of non-derivative moral and legal rights. Ascriptions of legal personhood to SARs, then, seems to lay regulatory foundations appropriate for the open-ended development of more sophisticated AI systems, to which would accrue more legal rights as their relationships with humans become more complex. At the same time, however, given their seeming intrinsic moral status, the production, sale and use of SARs endowed with general AI might appear to be radically immoral, more akin to sex trafficking than pornography. In other words, whilst the ascription of legal personhood is a core part of an appropriate legal framework for SARs embodying either *general* or *narrow* AI, the roles which general AI could appropriately occupy in the sphere of social robotics might turn out to be more limited than the roles appropriate to narrow AI.

³⁶¹ Turner (2019: 187-188). Bryson *et al.* (2017: 4.2.1) contend that ascriptions of legal personhood to AI systems would lead to ‘natural persons using an artificial person to shield themselves from the consequences of their conduct’. Turner (2019: 191) points out that this line of argument is not just speculative but moreover entails the absurd conclusion that we should abolish ‘limited liability’—and, thus, companies—altogether.

5.2. Conclusion

I have sought to make two principal contributions to the AI ethics literature in Part One of this dissertation. Firstly, I have articulated and defended a new argument for the legal proscription of robotic sexual violence, one which appeals to politically liberal principles which all free and equal citizens in a pluralistic society can reasonably accept. This line of argument is dialectically important because, until now, arguments for the legal proscription of robotic sexual violence have depended upon the invocation of moral conceptions of the good, whilst politically liberal principles have seemed to support the legal permissibility of robotic sexual violence.

Secondly, I have sought to shift the focus in the academic and public debate surrounding the ascription of legal personhood to autonomous robots. Legal personhood has been touted as a device which will assist us in developing an appropriate legal framework to regulate wrongs caused *by* autonomous robots. But I have argued in Part One of this dissertation that legal personhood may also assist us in developing an appropriate legal framework to regulate wrongs inflicted *on* autonomous robots. This is vitally important: what we owe to autonomous robots, both morally and legally, is a matter which we as private individuals and public legislators can no longer afford to ignore.

In sum, given the imminent emergence of autonomous robots in spheres both public and private, a legal framework for their regulation is urgently needed. Part One of this dissertation is my contribution to that wider project.

PART TWO

AGAINST GAPS: MORAL AND LEGAL RESPONSIBILITY FOR ACCIDENTS INVOLVING LETHAL AUTONOMOUS WEAPONS

Come writers and critics who prophesize with your pen

And keep your eyes wide, the chance won't come again

And don't speak too soon for the wheel's still in spin

And there's no tellin' who that it's namin'³⁶²

³⁶² Dylan (1964).

6. The Responsibility Gaps Thesis

(*What is it good for?*) Absolutely nothin³⁶³

6.1. Introduction

Part Two of this dissertation is about a thesis concerning responsibility which has emerged in the AI ethics literature. I call this the Responsibility Gaps Thesis.³⁶⁴ The Responsibility Gaps Thesis consists essentially in the claim that AI systems generate normatively problematic ‘responsibility gaps’, that is, some discrepancy between the (degree of) responsibility which accrues to any person or group of persons and the (degree of) responsibility which should be allocated to that person or group.³⁶⁵ The centrality of this thesis to the philosophical literature on AI ethics can hardly be overstated—indeed, it underpins some of the foremost academic and public debates in AI ethics of the last 20 years, including calls for, *inter alia*, a pre-emptive ban on the development of certain systems,³⁶⁶ ‘meaningful human control’ over autonomous AI,³⁶⁷ the legal personification of certain AI systems³⁶⁸ and explainable AI.³⁶⁹ But the Responsibility Gaps Thesis is false, at least as it is usually understood. In brief: chapters six and seven will establish that two prevalent interpretations of the Responsibility Gaps Thesis trade on,

³⁶³ Starr (1970).

³⁶⁴ The literature on ‘responsibility gaps’ involving machine learning systems in general is usually traced back to Matthias (2004); Sparrow (2007) is credited as the first to apply this idea to the production, sale and use of LAWs. As I discuss *infra* the AI ethics literature on ‘responsibility gaps’ may broadly be characterised in terms of different responses to the alleged challenge posed by Matthias and Sparrow. For critical examination of this literature including some engagement with the very idea of a ‘responsibility gap’: Santoni de Sio and Mecacci (2021); Veluwenkamp (2025). For instructive legal analysis of lethal accidents in ‘high-risk civil applications’—that is, ‘AI applications resulting in death or serious injury [where] death or injury is never an intended consequence, unlike in military applications where death may be intended, but mistakes may still occur’ (Grieman 2024: 5)—see Grieman (2024). Part Two of this dissertation focuses on lethal accidents involving death or serious injury in military applications.

³⁶⁵ Da Silva (2024).

³⁶⁶ See especially the Stop Killer Robots campaign and the series of articles published by Human Rights Watch (HRW 2012, 2018).

³⁶⁷ In 2013, Richard Moyes the head and managing partner of the UK-based NGO Article 36 coined the phrase ‘meaningful human control’ as a legal standard for the use of autonomous AI in exercises of lethal force (Article 36 2016; Roff and Moyes 2016; ICRC 2014: 45–48). See also: Ekelhof (2018); Amoroso and Tamburrini (2021b).

³⁶⁸ See Delvaux (2017) discussed *supra* in 1.1.

³⁶⁹ Mittelstadt *et al.* (2016).

respectively, false presumptions about the value of responsibility attributions *per se* or equivocation between different meanings of the term ‘responsibility’. Chapter eight considers a third, alternative interpretation of the Responsibility Gaps Thesis as a claim about the moral adequacy of legal responsibility mechanisms for autonomous AI in war. Moral, legal and political conversations regarding this third interpretation have so far focused on the moral adequacy of international criminal law.³⁷⁰ I propose a new focus. In essence, we need to recognise the Responsibility Gaps Thesis as a strong challenge to the moral adequacy of international private law in a world where fundamental decisions about human lives are increasingly taken by private actors, armed with learning algorithms. Chapter nine articulates and defends some reforms to the private law doctrines of vicarious liability and strict product liability which respond to this challenge; I argue *inter alia* for the ascription of legal responsibility to certain kinds of autonomous AI. A corollary is that those AI systems should be legal persons.

I focus on lethal accidents involving LAWs—viz., ‘weapons system[s] that, once activated, can select and engage targets without further intervention by a human operator’.³⁷¹ The ongoing proliferation of this kind of technology raises a host of normative questions.³⁷² Critics argue that LAWs threaten ‘global security and regional and international stability’³⁷³ through ‘increasing the risk of an emerging arms race’,³⁷⁴ lowering the threshold for the use of force³⁷⁵ and/or ‘proliferation, including to non-State actors’.³⁷⁶ LAWs may also be incapable of complying with various rules of international law³⁷⁷ or susceptible to malfunctioning,

³⁷⁰ Although see Crootof (2015, 2016, 2022), discussed *infra* at 8.2.2.2.

³⁷¹ This definition by the US Department of Defense (2012, 2023: 21) has emerged as the working definition in the academic and political debate surrounding LAWs. Still, there is some controversy over the precise definition, which I return to in 6.2 below.

³⁷² See generally: Sharkey (2012); Asaro (2020). For rejoinders: Meier (2019); Bellaby (2021).

³⁷³ UN (2023a).

³⁷⁴ *Ibid*; Heyns (2013: 31, [80]–[94]).

³⁷⁵ Singer (2009: 323); ICRC (2016: 16); UN (2023b, 2024: [12]).

³⁷⁶ UN (2023a).

³⁷⁷ See: Asaro (2013); Sharkey (2012); HRW (2012, 2013); Wagner (2012).

hacking and spoofing.³⁷⁸ Others, however, point out that LAWs can do ‘the so-called dirty, dull and dangerous work’ of warfare,³⁷⁹ thereby protecting the lives of a nation’s or a state’s own soldiers,³⁸⁰ and may ‘help to better protect civilians in conflict in certain circumstances’.³⁸¹ It is further argued that LAWs will be immune to human limitations such as fatigue³⁸² and will ‘not act out of revenge, panic, anger, spite, prejudice or fear’,³⁸³ reactions which may be to the detriment of compliance with just war principles.³⁸⁴ Some propose that LAWs may be equipped with the ‘potential capability of independently and objectively monitoring ethical behavior in the battlefield by all parties and reporting infractions that might be observed’ in order to drive human compliance with international legal rules through augmenting accountability for wrongdoing.³⁸⁵ Others believe that LAWs will improve the accuracy of lethal strikes at a distance, which would reduce not just the number of human soldiers exposed to physically and psychologically hazardous environments³⁸⁶ but also collateral damage.

This dissertation explores different terrain.³⁸⁷ Initially, in the rest of chapter six, I set out what LAWs are and one standard argument for the Responsibility Gaps Thesis.³⁸⁸ The argument is, in essence, that no one is morally responsible for certain kinds of harms caused by autonomous systems like LAWs but *someone ought to be*. It is on this latter proposition that chapter six focuses. I submit that it is false, so one standard argument for the Responsibility Gaps Thesis is unsound. This prompts me to consider whether there is an alternative

³⁷⁸ Heyns (2013: 18); Singer (2009: 261-263).

³⁷⁹ Heyns (2013: 10, [51]).

³⁸⁰ US Department of Defense (2018); Marchant *et al.* (2011).

³⁸¹ UN (2023b); Arkin (2009, 2014); Marchant *et al.* (2011).

³⁸² Lin *et al.* (2008); Krishnan (2009: 39-42).

³⁸³ Heyns (2013: 10, [54]); Marchant *et al.* (2011).

³⁸⁴ Walzer (1977: 251).

³⁸⁵ Arkin (2014: 9).

³⁸⁶ Singer (2009: 30-37).

³⁸⁷ Following Eggert (2023), I therefore situate this discussion of autonomised harming in broadly non-consequentialist ethics. For a general defence of this kind of position: Kamm (2007).

³⁸⁸ Matthias (2004); Sparrow (2007).

interpretation of the Responsibility Gaps Thesis which might explain its centrality in the AI ethics discourse.

Chapter seven takes up an alternative interpretation of the Responsibility Gaps Thesis espoused by authors including Da Silva and Köhler and others.³⁸⁹ The idea is that there is some kind of gap between what our best theories of responsibility say about certain kinds of harms involving LAWs and what our intuitions say about those kinds of harms. In other words, the claim goes, there are cases of harms involving LAWs which are counterexamples to our best theories of responsibility and give us good reason to revise them. That claim is also false. I begin in chapter seven by distinguishing two different concepts which people might mean when they use the word ‘responsible’ to describe an agent’s relationship to some outcome. On one hand, ‘responsible’ might mean that it is apt to demand that the agent justify, explain or generally ‘answer for’ the role which they played in bringing the outcome about; on the other hand, ‘responsible’ might mean that it is apt to blame, punish or sanction the agent for what has happened. I take each of these meanings of ‘responsibility’ in turn and consider what our best theories of each kind of responsibility say about cases involving purported ‘responsibility gaps’. Then, I ask whether there is any discrepancy between what our best theories say about those cases and our intuitions. I say that there are no discrepancies and, therefore, no gaps.

Chapter eight considers a third and final kind of interpretation of the Responsibility Gaps Thesis, namely that there is a discrepancy between the (degree of) legal responsibility and the (degree of) moral responsibility which is attributable in cases of purported ‘responsibility gaps’. On one hand, some have suggested that there is a gap between what is in fact criminalised under international law and who it is intrinsically or instrumentally good to

³⁸⁹ Da Silva (2024); Köhler *et al.* (2017).

punish.³⁹⁰ To show that there is a gap of this kind, we need to establish at least two propositions. First, that existing criminal law mechanisms do not in fact apply to at least some cases involving LAWs. Second, that it is intrinsically or instrumentally good to punish in those cases. I argue that there is no gap between what is in fact criminalised and who it is intrinsically or instrumentally good to punish in cases involving LAWs.

On the other hand, there might be a gap between the kinds of remedies which are enforceable under international private law as is currently stands and the reparatory moral duties which people have in cases of purported ‘responsibility gaps’. To show that there is a gap of this kind, we need to show, first, that people do in fact have reparatory moral duties of the kind which should be legally enforceable in certain cases involving LAWs and, second, that existing private law mechanisms do not apply to those cases. It is here that I identify a novel, true interpretation of the Responsibility Gaps Thesis: I argue that there is a gap between the reparatory moral duties which people have in certain lethal accidents involving LAWs and the kind of remedies available under international private law. This prompts the question: how should we *change* the law to close that gap?

Chapter nine contains novel and timely proposals for law reform informed by the preceding arguments. I submit that international private law should reflect the fact that those persons and groups of persons which produce, sell and use LAWs are participating in enterprises which create certain characteristic risks to others. Those producers, sellers and users should therefore be legally responsible for compensating those who are harmed by LAWs, even in the absence of fault. There are two principal mechanisms for redistributing risk in cases like this: vicarious liability and strict product liability. Chapter nine articulates and defends certain targeted amendments to these legal mechanisms to promote justice and fairness following

³⁹⁰ Danaher (2016a); De Vries (2023). For criticism of the retributivist view that punishment is justified on the basis of moral desert: Tadros (2011: ch.4).

lethal accidents involving LAWs. First, I argue for the imposition of vicarious liability for lethal accidents on the military enterprises which deploy LAWs to ensure the legal expression and enforcement of those reparatory moral duties which attach to enterprises that *use* LAWs as quasi-employees. A corollary which I prove is that LAWs should be legal persons bearing legal duties not to target civilians. Second, I argue for targeted revisions of existing strict product liability mechanisms to ensure the legal expression and enforcement of those reparatory moral duties which are owed by enterprises that *produce* and *sell* LAWs. I suggest that the result is a legal framework fit for purpose in a world where AI systems cause lethal accidents as *quasi-employees* of some enterprises and *products* of others. This completes my answer to the Responsibility Gaps Thesis.

6.2. What are LAWs?

Following publication of the Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions, Christof Heyns,³⁹¹ State parties to the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons [...] formed the Group of Governmental Experts on LAWs.³⁹² Notwithstanding a decade of meetings,³⁹³ there is no commonly agreed definition of LAWs.³⁹⁴ I will assume that LAWs display the following three features based on the commonalities evinced by these discussions.

6.2.1. Lethal independence

I assume that LAWs exhibit autonomy in ‘critical functions’ such as selecting and engaging human targets.³⁹⁵ Indeed, as discussed in Part One, it is common to distinguish three levels of human control over autonomous systems: ‘human-in-the-loop’ weapons select targets and

³⁹¹ Heyns (2013).

³⁹² Bo *et al.* (2022); Solovyeva and Hynek (2023); Autonomous Weapons (2025a).

³⁹³ Autonomous Weapons (2025a, 2025b).

³⁹⁴ For an overview of different characterisations of these systems: UN GGE (2023).

³⁹⁵ ICRC (2016); Meier (2019: 293).

deliver force only with human input; ‘human-on-the-loop’ ones select targets and deliver force without human input, although a human monitor can override or intervene; ‘human-out-of-the-loop’ weapons select targets and deliver force with neither human input nor a human monitor with the ability to intervene.³⁹⁶ Only human-on-the-loop and human-out-of-the-loop systems count as autonomous for the purposes of the LAWs literature. This is why most current unmanned aerial vehicles, or drones, are not LAWs.³⁹⁷

LAWs may therefore be distinguished from many autonomous weapons currently in use.³⁹⁸ For example, the Counter-Rockets, Artillery and Mortar (C-RAM) is a set of systems used to automatically detect and destroy incoming rockets, artillery, and mortar rounds in the air, before they hit the ground.³⁹⁹ The US naval Phalanx Weapon System for Aegis-class cruisers ‘automatically detects, evaluates, tracks, engages and performs kill assessment against anti-ship missiles and high-speed aircraft threats’.⁴⁰⁰ There is also a land-based Phalanx Weapon System known as Centurion C-RAM. Other C-RAM systems include: Israel’s Iron Dome anti-missile system, featuring multiple-target tracking and self-guided missile interceptors, and Germany’s MANTIS and Skynex air defence systems.⁴⁰¹ Israel’s Harpy is a so-called ‘fire-and-forget’ autonomous weapon system designed to detect, attack and destroy radar emitters;⁴⁰² the UK’s Taranis is a jet-propelled unmanned air system which ‘is capable of undertaking sustained surveillance, marking targets, gathering intelligence, deterring adversaries and carrying out strikes in hostile territory’⁴⁰³ although, as UN Special Rapporteur

³⁹⁶ Horowitz and Scharre (2015); Burri (2017); HRW (2015: n.3); Wagner (2016).

³⁹⁷ Mayer (2015: 766).

³⁹⁸ Seixas-Nuñez (2024: 195) critiques the US Department of Defense’ definition of an autonomous weapon system on the basis that it does not clearly rule out those existing systems (e.g., the Phalanx Weapon System or Harpy) which do not display lethal independence. This dissertation focuses on systems displaying lethal independence since these generate the most interesting normative and conceptual challenges.

³⁹⁹ IAMD ‘C-RAM’.

⁴⁰⁰ Raytheon ‘Phalanx’.

⁴⁰¹ Bellaby (2021); Meier (2019).

⁴⁰² IAI ‘HARPY’.

⁴⁰³ BAE ‘Taranis’.

Heyns clarifies, it can only engage with a target when authorized by mission command.⁴⁰⁴ By contrast, once we specify that autonomous weapons systems become LAWS only if they exhibit autonomy in critical functions, many of these existing applications are ruled out.

A caveat: the term ‘autonomy’ enjoys a broad usage in the popular, academic and techno-military media which overlaps but lacks coextension with the definition applied here. For example, in June 2025, Ukraine’s security service (SBU) carried out ‘Operation Spiderweb’ against four air bases in Russia. As the Financial Times reported: ‘AI-powered autonomous systems allowed the drones to keep flying ‘along a pre-planned route’ even after losing signal.’⁴⁰⁵ The SBU reported that ‘[a]fter approaching and contacting a specifically designated target, the warhead was automatically triggered.’⁴⁰⁶ Although it is difficult to determine without a detailed specification of its capabilities, it appears that the system lacks lethal independence because the route was ‘pre-planned’ and the targets ‘specifically designated’. Operation Spiderweb does not, then, evince a system capable of *selecting* and engaging targets, even after losing signal.

6.2.2. Adaptability

One system with a degree of lethal independence in target selection and, possibly, engagement is the Samsung Techwin Sentry Guard Robot (SGR-A1), which is deployed in the demilitarised zone between North and South Korea. For use in the demilitarised zone, the system does not need to distinguish friend from foe and is capable of firing automatically when someone crosses the line although it can, allegedly, detect a ‘hands-up’ sign of surrender in order to hold fire. Although the ultimate decision about target engagement would normally be made by a human, the robot has an ‘automatic mode’ in which it can ‘make the decision’.⁴⁰⁷ This is somewhat

⁴⁰⁴ Heyns (2013: 9).

⁴⁰⁵ Deprez (2025).

⁴⁰⁶ *Ibid.*

⁴⁰⁷ Global Security ‘Military’.

similar to the automatic target engagement possible through the use of landmines. Nevertheless, neither the SGR-A1 nor landmines would qualify as LAWs because it is standard in the literature to require that some level of *adaptability*, in addition to lethal independence, is required.⁴⁰⁸

One way to think about the relationship between autonomy and adaptability is as follows:

‘[A]utonomy is a function of [...] three variables [...]: independence, adaptability, and discretion. A system is autonomous when it acts with sufficiently infrequent operator interaction, when it is able to function successfully in unfamiliar environments, and when it achieves mission objectives with a high level of assertiveness [...] As a result, ‘[l]ike intelligence and consciousness,’ autonomy is best conceived of as existing on a spectrum. Some machine systems would clearly lie on the ‘automated’ end, like a robot welding doors in a Ford Motors plant. Other systems might be closer to autonomous—think of a futuristic drone able to seek, identify, and kill an enemy without any human interaction.’⁴⁰⁹

This goes to the heart of the difference between automated weapons and autonomous weapons: automated systems ‘operate within a structured and predictable environment’ whilst ‘[a]utonomous systems can function in an open environment, under unstructured and dynamic circumstances’.⁴¹⁰ As such, although high in lethal independence, the SGR-A1 and landmines appear to be quite low in autonomy because they lack adaptability and, as we will see next, discretion.

⁴⁰⁸ See: Seixas-Núñez (2022a: 429-439; 2022b: 82-89; 2024).

⁴⁰⁹ Marra and McNeil (2012: 1155).

⁴¹⁰ Heyns (2013: 8); c.f. Chengeta (2016).

6.2.3. Discretion

It is important to clarify at the outset the nature of the AI at issue in LAWs since this determines the sense in which LAWs have ‘discretion’ in the selection and engagement of targets. There are two basic understandings of discretion:⁴¹¹ the US Department of Defense’ function-oriented conception, according to which LAWs ‘once activated, can select and engage targets without further intervention by a human operator’,⁴¹² and the UK Ministry of Defence’ intent-oriented view that LAWs should be ‘capable of understanding higher level intent and direction’.⁴¹³ The intent-oriented conception seems to require something close to that elusive *general* AI, which selects its own goals, as opposed to *narrow* AI, which involves the skilful pursuit of a pre-set goal and is manifest in present technology.⁴¹⁴ I adopt the US Department of Defense’ conception because, since they already exist, it is to accommodate LAWs of *this* description that we most urgently need a robust normative framework.⁴¹⁵ In narrow AI, ‘the system is designed to complete engagements within a timeframe and geographic area, as well as other applicable environmental and operational parameters, consistent with the commander and operator intentions’.⁴¹⁶ In order to achieve this level of discretion, the system must be ‘able to change its own internal states to achieve a given goal, or a set of goals, within its dynamic operating environment and without the direct intervention of another agent’.⁴¹⁷ In other words, the system must implement some kind of machine learning.⁴¹⁸

⁴¹¹ Amoroso and Tamburrini (2021a: 21-22).

⁴¹² US Department of Defense (2012).

⁴¹³ Ministry of Defence (2011: 14; 2017: 21).

⁴¹⁴ Amoroso and Tamburrini (2021a: 21-22).

⁴¹⁵ LAWs were deployed in Libya and possibly Palestine in 2021 (UN Libya Report 2021; Buhbut 2021). By 2035, the US DoD envisions the integration of LAWs into the force (JCRAS Report 2016).

⁴¹⁶ US Department of Defense (2023: §4, 4.1).

⁴¹⁷ Taddeo and Blanchard (2022: 15).

⁴¹⁸ For existing applications of machine learning in autonomous weapons systems: Trabucco (2025: 542-544). This includes the Ukrainian-made Saker Scout drone which ‘use[s] machine learning for object recognition to identify targets and perform visual navigation in the event of Global Positioning System jamming [...] the drone uses machine learning models for object recognition to identify and engage preauthorized targets without human intervention.’ (*ibid*: 542-543).

Broadly, ‘a machine learns whenever it changes its structure, program, or data (based on its inputs or in response to external information) in such a manner that its expected future performance improves’.⁴¹⁹ In particular, LAWs are likely to implement either unsupervised or weakly supervised (i.e. reinforcement) learning. In unsupervised learning, the algorithm is given training data with no labels or feedback.⁴²⁰ Unsupervised systems function by grouping the data into clusters of similar features. The programmers may not know of these patterns in the data beforehand, which is why unsupervised systems are often touted as a mode of ‘discover[ing]’ knowledge’.⁴²¹ In weakly supervised learning, the system learns to take ‘actions’ in a dynamic environment so as to maximise the cumulative reward. In more detail, a basic reinforcement learning system interacts with its environment in discrete time steps. At each time t , the system receives the current state S_t and reward R_t . It then chooses an action A_t from the set of available actions, which is then sent to the environment. The environment moves to a new state S_{t+1} and the reward associated with the ‘transition’ (S_t, A_t, S_{t+1}) is determined. The goal of the system is to learn a policy, that is, a mapping of states and actions onto a measure space, which maximises the expected cumulative reward. As Turner puts it, ‘[t]he program is not told which actions to take, but instead has to discover which actions yield the most reward through an iterative process’.⁴²²

Applied in war, LAWs will therefore display a level of complexity in responding to stimuli not met by automatic weapons such as landmines, South Korea’s SGR-A1 sentry bot or Israel’s Iron Dome anti-missile battery.⁴²³ In the literature, it is this feature which is standardly taken to differentiate LAWs from other weapons. As Kiener succinctly explains,

⁴¹⁹ Nilsson (2015: 1).

⁴²⁰ This is distinguished from (strongly) supervised learning, where the algorithm is given labelled training data, that is, training data which contains the ‘correct answer’ (Turner 2019: 71).

⁴²¹ Boden (2016: 47).

⁴²² Turner (2019: 73); c.f. Sutton and Barto (1998: 4).

⁴²³ Bellaby (2021); Meier (2019).

‘[t]he best [AI] is often the least controllable and the least transparent’.⁴²⁴ As regards control, this is because processes become too fast to be monitored in real time.⁴²⁵ As regards transparency, this is because the best AI relies on technological bases like deep neural networks which create black-box AIs with impenetrably complex algorithms.⁴²⁶ The upshot is that a tried-and-tested AI can go wrong in seemingly uncontrollable and unforeseeable ways. So, LAWs might be overwhelmingly successful at distinguishing military targets from civilian objects but for the wrong reasons, such that they go wrong in unexpected cases.⁴²⁷

6.3. The Responsibility Gaps Thesis

Consider the following example.⁴²⁸ To make matters stark, this dissertation makes the idealising assumptions that the LAW is tried-and-tested, more reliable than the average human pilot at distinguishing between civilian objects and military targets and deployed in a war which satisfies the *jus ad bellum* and *jus in bello* conditions for just war, whatever those might be.⁴²⁹

Massacre: A LAW bombs a crowd of civilians. These civilians posed no threat to friendly forces or non-combatants. This is not a mechanical malfunction and no human intended or foresaw it; rather, it results from the weapon’s algorithmic selection-and-engagement process. It is clear that ‘had a human being committed the act, they would [...] be charged with a war crime.’⁴³⁰

⁴²⁴ Kiener (2022: 575).

⁴²⁵ Coeckelbergh (2020); Sparrow (2007).

⁴²⁶ Bathaee 2018; Wang *et al.* 2019.

⁴²⁷ C.f. *supra* n.352.

⁴²⁸ Sparrow (2007).

⁴²⁹ The classic text is Walzer (1977); c.f. Rawls (1999: §§13-14). I gloss over two controversies in the just war literature. The first concerns whether there is an additional, distinct category of *jus post bellum* conditions (Bellamy 2008). The second concerns the appropriate distinction between the *jus ad bellum* and *jus in bello* categories. Since the resolution of these disputes is orthogonal to my arguments, I take no stand on either issue but refer the reader to Frowe (2023).

⁴³⁰ Sparrow (2007: 66). I assume that ‘the act’ denotes a type not a token and, as such, there is no assumption that acts are preserved under change of causal agent here. Thanks to David Enoch for pressing me on this point.

Lots of authors seem to think that this kind of case exemplifies the Responsibility Gaps Thesis (6.3.1). But what does the notion of a ‘gap’ in responsibility really mean? One common view seems to be that no-one is responsible for the outcome in cases like Massacre but *somebody ought to be*. I suggest that this presupposes that there is some value, all else being equal, in someone being responsible for the harm in cases like Massacre. I show in 6.3.2 and 6.3.3 that that presumption is false. 6.3.4 outlines two alternative interpretations of the Responsibility Gaps Thesis which I examine in chapters seven and eight.

6.3.1. Popular perspectives on the Responsibility Gaps Thesis

The discussion of the Responsibility Gaps Thesis is often traced to Matthias’ argument that the development of ‘learning automata’—that is, autonomous systems equipped with machine learning technology—means that:

‘there is an increasing class of machine actions, where the traditional ways of responsibility ascription are not compatible with our sense of justice and the moral framework of society because nobody has enough control over the machine’s actions to be able to assume the responsibility for them. These cases constitute what we will call the responsibility gap.’⁴³¹

This, Matthias claims, introduces a dilemma: either we terminate the development of these systems and forego their conveniences, or we embrace their development and resign ourselves to ‘responsibility gaps’. It is implicit in this dilemmatic framing that there is something at least *prima facie* problematic about ‘responsibility gaps’.

At this juncture, the literature divides into what Tigard has described as the ‘techno-pessimists’ and the ‘techno-optimists’.⁴³² Techno-pessimists believe that ‘technology is posing

⁴³¹ Matthias (2004: 177).

⁴³² Tigard (2020).

a unique and potentially insurmountable obstacle to our usual sense of morality and law’ such that ‘we must drastically scale back or altogether cease our deployment of AI systems’.⁴³³ Techno-optimists, by contrast, believe that ‘the gap can be bridged’ through technological or legal innovations.⁴³⁴

Techno-pessimists have argued *inter alia* for a pre-emptive ban on the production, sale and/or use of LAWs on the basis of the Responsibility Gaps Thesis. Indeed, Sparrow thinks that the Responsibility Gaps Thesis represents a deep challenge to the moral and legal permissibility of the production, sale and use of LAWs altogether. He writes:

‘[i]t is a necessary condition for fighting a just war [...] that someone can justly be held responsible for deaths that occur in the course of the war. As this condition cannot be met in relation to deaths caused by an autonomous weapon system it would therefore be unethical to deploy such systems in warfare.’⁴³⁵

Similarly, Human Rights Watch has argued that ‘states should eliminate [the] accountability gap [associated with the use of LAWs] by adopting an international ban on fully autonomous weapons.’⁴³⁶

Techno-optimists have argued for novel legal and technical solutions to remedy the practical challenges posed by LAWs.⁴³⁷ Proposed legal solutions include the introduction of a new kind of legal personhood, namely ‘e-personhood’, to enable the attribution of legal responsibility to LAWs themselves;⁴³⁸ the introduction of a new legal standard of ‘meaningful

⁴³³ Tigard (2020: 590). See, e.g., Sharkey (2010); Asaro (2012); Char *et al.* (2018); Danaher (2016b).

⁴³⁴ See, e.g., Marino and Tamburrini (2006); Hanson (2009); Rahwan (2018); Nyholm (2018).

⁴³⁵ Sparrow (2007: 62).

⁴³⁶ HRW (2015).

⁴³⁷ The distinction between legal and technical solutionism appears in Santoni de Sio and Mecacci (2021).

⁴³⁸ *Supra* n.368, 1.1.

human control’;⁴³⁹ and the introduction of new kinds of criminal⁴⁴⁰ and civil⁴⁴¹ liability. Among the technical solutions which have been touted, explainable AI may mitigate the risk of opaque, black-box AI.⁴⁴² My purpose here is merely to adumbrate the landscape of responses which the Responsibility Gaps Thesis has prompted, rather than to offer any sustained critical examination of the different positions. Notice, though, that what techno-optimists and techno-pessimists appear to agree upon is the Responsibility Gaps Thesis itself.

An alternative third way is sketched out by so-called ‘deflationists’,⁴⁴³ namely those who deny the Responsibility Gaps Thesis. For instance, some have argued that there is no ‘gap’, that is, no discrepancy between the (degree of) responsibility which accrues to any person or group of persons in cases like Massacre and the (degree of) responsibility which ‘should’ be allocated to a person or group of persons in such cases.⁴⁴⁴ Others have argued that, even if there were a ‘gap’ of some kind, it would not be a new or distinctive⁴⁴⁵ problem—or even a problem at all⁴⁴⁶—for normative theory. My contribution in Part Two of this dissertation is to offer the first sustained critical examination of what people really mean when they claim that cases like Massacre involve (or do not involve) a ‘responsibility gap’. There are at least three distinct kinds of normative claims being made, and only one of these has any philosophical merit. My approach will have notes of techno-optimism and deflationism, but not much techno-pessimism. To begin with, let us look at the most common kind of normative claim which people seem to be making when they state the Responsibility Gaps Thesis.

⁴³⁹ See references *supra*: n.367.

⁴⁴⁰ Burri (2018); Heyns (2013); Corn (2014).

⁴⁴¹ Crotoft (2016).

⁴⁴² See: *supra* n.369.

⁴⁴³ This term derives from Santoni de Sio and Mecacci (2021) although I use it to embrace a wider range of authors than they do.

⁴⁴⁴ See, e.g., Himmelreich (2019); Robillard (2018) and, for conceptual revision of the concept of a ‘gap’, Veluwenkamp (2025).

⁴⁴⁵ Simpson and Müller (2016); Demirtas (2025); Tigard (2020).

⁴⁴⁶ Danaher (2022); Munch *et al.* (2023).

6.3.2. The value of responsibility per se?

The most straightforward reading of authors like Sparrow is that the Responsibility Gaps Thesis is prompted by cases like Massacre because, in such cases, no one is responsible for the outcome but *someone ought to be*. This seems to presuppose that it is better, other things being equal, if someone is responsible for the harm than if no one is. But is that so? This sub-section considers whether it is *intrinsically* morally valuable for someone to be responsible for a harm. By considering three different perspectives on an injury—namely, the victim’s, the perpetrator’s and that of a godlike judge of the overall moral value of a situation—I suggest that it might, on the contrary, be better if no one is responsible for some harm than if someone is. The next sub-section considers whether there are *instrumental* moral reasons to think that it is better if someone is responsible for harm than not.

To begin with, let us consider the perspective of a victim. Consider the following two cases:

Victim of psychotic agent: In world 1, I wake up from a coma and am informed that my injury was caused by the actions of a psychotic agent who has since died.

Victim of responsible agent: In world 2, I wake up from a coma and am informed that my injury was caused by the actions of an agent who freely and knowingly decided to hurt me but has since died. All else is the same as in world 1.

I would rather wake up in world 1. The reason for this is simple: worlds 1 and 2 are otherwise identical but in world 2 someone is responsible for my injury and in world 1 no one is; it is better from the victim’s perspective to be injured non-responsibly. If you are unpersuaded, let us emphasise that all else is equal between worlds 1 and 2: given that the agent of my injury has died in both worlds, there is no question of getting my retributive instincts satisfied through punishment and no reason in either case for me to fear further injury; let us also assume that I

am not getting any more compensation for my injuries in world 1 or world 2. With instrumental considerations about responsibility therefore set aside, I submit that we prefer to be hurt by agents who, in some sense, did not freely and knowingly decide to hurt us.

What about the perpetrator's perspective? Let us assume, for the sake of argument, that there is an afterlife and you are going to be treated just the same in the afterlife regardless of how well or poorly you behaved whilst alive. Consider, next, the following two cases:

Psychotic agent: In Afterlife 1, I wake up with no memory of my life. I am informed by a helpful angel that just before dying I had a psychotic episode during which I injured someone.

Responsible agent: In Afterlife 2, I wake up with no memory of my life. I am informed by a helpful angel that just before dying I freely and knowingly injured someone. All else is the same as in Afterlife 1.

I would rather wake up in Afterlife 1. The reason is, again, simple: it is better from the perpetrator's perspective not to bear responsibility for harm that she has caused.

I have suggested that Universe 1 (containing World 1 and Afterlife 1) is better for the victim *and* the perpetrator than Universe 2 (containing World 2 and Afterlife 2). It is plausible, given this finding, that Universe 1 is better overall than Universe 2. In other words, it is better for no one to be responsible for a given injury than for someone to be responsible. Now, let me persuade you that it is not better, other things being equal, if someone is responsible for harm like that in Massacre than if no one is. Recall that for Sparrow part of the reasoning is that '[i]t is a minimal expression of respect due to our enemy'⁴⁴⁷ and, indeed, 'an important requirement

⁴⁴⁷ Sparrow (2007: 67).

of *jus in bello*⁴⁴⁸ that ‘someone should accept responsibility, or be capable of being held responsible, for the decision to take their life.’⁴⁴⁹ But consider:

Psychotic Soldier. In world 3, the outcome is the same as in Massacre except that in world 3 the bombing is carried out by a soldier flying the drone remotely who, in the midst of a psychotic episode, mistakenly thinks that the attack is justified.

Responsible Soldier. In world 4, everything is the same as in world 3 but this time the bombing is carried out by a soldier flying the drone remotely who freely and knowingly decides to kill the civilians.

Let us assume that in both cases this episode was entirely unforeseeable by the soldier’s commanders who had taken reasonable care in her training. Let us also assume, as is plausible, that in Psychotic Soldier, the perpetrator of the massacre is a fully non-responsible agent given that she does not satisfy the standard epistemic or control conditions for responsible agency (e.g., she does not know the nature of her actions). If Sparrow is right that ‘[i]t is a minimal expression of respect due to our enemy’⁴⁵⁰ that ‘someone should accept responsibility, or be capable of being held responsible, for the decision to take their life’,⁴⁵¹ then Responsible Soldier is a case where there is more respect for the enemy than Psychotic Soldier. But that seems absurd: why would a deliberate massacre be better or more respectful than a non-responsible one? I suggest that, setting aside instrumental reasons (relating to, say, compensation) why we might wish to find a responsible agent, there is no disvalue at the level of the second-order moral reasons applicable to a given situation if no one bears responsibility for the outcome.

⁴⁴⁸ *Ibid.*

⁴⁴⁹ *Ibid.*

⁴⁵⁰ *Ibid.*

⁴⁵¹ *Ibid.*

6.3.3. The instrumental value of responsibility?

Let us now consider what kinds of instrumental moral reasons we might have for preferring a world in which someone is responsible for a given harm to a world in which no one is responsible for that same harm. Indeed, some authors have suggested that LAWs create some kind of ‘responsibility gap’ because there is an instrumental value in having someone to ‘hold responsible’. This might be in order to have someone to punish,⁴⁵² someone to land with special duties of compensation⁴⁵³ or incentives to take reasonable care.⁴⁵⁴ I argue that none of these reasons are convincing.

As to the issue of punishment, Danaher has argued that cases like Massacre increase the risk of two types of negative consequence. First, he believes that there is a risk of moral scapegoating. This is because, in cases like Massacre, ‘there is a deep human desire to find appropriate targets of retributive blame’ but if ‘none really exist, then there is a danger that people will try to fulfil that desire in inappropriate ways’ by, say, punishing manufacturers, programmers or military officers.⁴⁵⁵ If they cannot be ‘held responsible’, this amounts to punishing the innocent. Second, Danaher fears that cases like Massacre might undermine respect for the rule of law. This appeals to the idea defended by some legal theorists that the rule of law can be undermined if punishment practices within legal systems fail to align with the reasonably fixed intuitions which people have about what kinds of behaviours or outcomes call for retribution.⁴⁵⁶ As such:

‘If people feel that *someone* deserves retributive blame for the harmful acts of robots, but our legal and moral systems are incapable of finding anyone, you will have a

⁴⁵² Danaher (2016a); c.f. Santoni de Sio and Mecacci (2021). There are of course other problems with a purely instrumental justification of punishment (Tomlin 2017; *infra* n.623). I defer discussion of intrinsic and instrumental justifications for punishment in cases like Massacre until **8.1.2**.

⁴⁵³ Santoni de Sio and Mecacci (2021).

⁴⁵⁴ *Ibid.*

⁴⁵⁵ Danaher (2016a: 307).

⁴⁵⁶ See: Robinson (2013); Robinson and Kurzban (2007); Robinson *et al.* (2007).

situation in which intuitive judgments are out of line with legal practice. This could begin to erode respect for the rule of law.’⁴⁵⁷

I am not persuaded by Danaher’s arguments because his two concerns seem to complete their own *reductio*. On one hand, we should not punish just to satisfy retributive instincts (since this would amount to moral scapegoating); on the other, we should aim to satisfy retributive instincts in our legal practices (since to fail to do so would allegedly undermine the rule of law). A concern for justice should make us prioritise the mitigation of any substantiated risk of moral scapegoating over tendentious arguments for the democratisation of sentencing decisions. But Danaher offers no evidence to support the speculation that cases like Massacre would increase that kind of risk.

As to the issue of compensation, Santoni de Sio and Mecacci suggest that ‘attributions of culpability have the function to compensate the victims—symbolically, or even materially, typically in the case of compensations to plaintiffs in civil litigations’.⁴⁵⁸ As such, if no one bears responsibility for harms such as Massacre, ‘victims of unjust harm will be less likely to receive compensation’.⁴⁵⁹ There are two reasons why this is not a persuasive reason to endorse the Responsibility Gaps Thesis. Firstly, given the stipulated assumptions, LAWs would likely be *better* than a human alternative at reliably distinguishing between civilian and military targets and ensuring the proportionality of strikes. So, there would be fewer victims of unjust harm overall. Of course, of those unjust harms that occur, a greater proportion might be caused through the kind of no-fault lethal accident involved in Massacre. But it would seem bizarre to claim that a world where there are more victims of unjust harm overall but a somewhat greater proportion of those victims receive compensation is morally better than a world where there

⁴⁵⁷ Danaher (2016a: 308).

⁴⁵⁸ Santoni de Sio and Mecacci (2021: 1062).

⁴⁵⁹ *Ibid*: 1063.

are fewer victims, especially given certain concerns about the moral adequacy—both in principle and in practice—of existing reparatory duties under international law.⁴⁶⁰ Secondly, and independently of that first point, I will argue in chapter eight that the developers of the LAW bear substantial reparatory moral duties in cases like Massacre and in chapter nine I show how the law should be set up to extract the appropriate reparations from them.

Finally, responsibility attributions license a variety of responses which assist with incentivising safer weapons. Indeed, if no one bears responsibility for Massacre, that fact might itself embolden leaders to wage more unjust wars or to take less care in the development of safe and reliable LAWs.⁴⁶¹ For one thing, though, this is not a reasonable inference on the facts. I have already stipulated that the LAW is tried-and-tested, better than the average human pilot at distinguishing civilians from military targets and deployed in a just war, whatever that might be. Indeed, as I discuss in the next chapter, everyone in Massacre is *a fortiori* taking all reasonable care. Let us not speculate, then, in the idealised example, about whether the attribution of responsibility or not in Massacre is going to have any particular consequences for the incentivisation of safe weapons or for the future conduct of war in cases which are not relevantly analogous. For another thing, as we will see in the next chapters, the decision to use a LAW with the intention of causing wrongful harm, or recklessness regarding such an eventuality, would not abrogate responsibility for the outcome: in brief, my responsibility in such a case would be traceable to my earlier decision to set in motion an unpredictable, uncontrollable or unstoppable war machine.

In sum, I have argued that it is not better, all things considered and all else equal, for someone to be responsible for the harm in Massacre than not. So: the claim that—in broad terms—*no one is responsible for the harms in Massacre but someone ought to be* is false. That

⁴⁶⁰ Crootof (2016, 2022); Eggert (2022a, 2022b).

⁴⁶¹ *Ibid.*

straightforward interpretation of the Responsibility Gaps Thesis is therefore to be rejected. But people still seem to think that there is a ‘responsibility gap’ in cases like Massacre. What might that gap involve?

6.3.4. Alternative interpretations of the Responsibility Gaps Thesis

We have seen that the most straightforward interpretation of the Responsibility Gaps Thesis—namely, that it is better, all else equal, if someone is responsible for the harm in cases like Massacre than if no one is—is false. There appear to be two alternative interpretations of the Responsibility Gaps Thesis in the recent literature. On one hand, some people mean that purported ‘responsibility gaps’ cases like Massacre are ones which evince a ‘gap’ between the (degree of) responsibility which accrues to any person or group of persons according to our best theories of responsibility and the (degree of) responsibility which accrues to any person or group according to our intuitions.⁴⁶² In other words, alleged ‘responsibility gaps’ cases such as Massacre purport to be counterexamples to our best theories of responsibility and, as such, reasons to revise those theories. It is to the examination of this kind of interpretation of the Responsibility Gaps Thesis that I turn in chapter seven, where I show that our best theories of responsibility, contrary to popular belief, actually get the intuitively right answers in cases such as Massacre. Such cases are not, therefore, counterexamples to our best theories of responsibility and offer no reason to revise those theories.

On the other hand, some people have interpreted the Responsibility Gaps Thesis as a claim about the existence of a discrepancy between *legal* responsibility and *moral* responsibility, that is, as a gap between what is in fact legally sanctioned and what should be sanctioned from a moral perspective.⁴⁶³ In chapter eight, I show that there is no discrepancy (at

⁴⁶² See, e.g., Da Silva (2024); Köhler *et al.* (2017). In the collective responsibility context, Collins (2019: 943) frames the issue as ‘holes’ in ‘situations in which we are unable to attribute all the responsibility we might pre-reflectively want to attribute [...]’.

⁴⁶³ See, e.g., Glavaničová and Pascucci (2022).

least not in principle) between the criminal responsibility attributable to a person or group of persons in cases like Massacre and the kinds of sanctions which are morally warranted; however, I will suggest that there is a discrepancy between the reparatory moral duties which people have in cases like Massacre and what can be achieved under international private law as it currently stands. Chapter nine articulates and defends a solution to that one kind of ‘responsibility gap’ which I identify in this dissertation as meritorious. But before that, let us look more closely at the claim that cases like Massacre might constitute some kind of counterexample to our best theories of responsibility.

7. Answerability and blameworthiness

*The language of responsibility is partly to blame.*⁴⁶⁴

This chapter critically examines an alternative interpretation of the Responsibility Gaps Thesis which some authors espouse.⁴⁶⁵ On this reading, purported ‘responsibility gaps’ cases like Massacre are ones which evince some kind of inadequacy in our best theories of responsibility. For example, Da Silva defines a ‘responsibility gap’ as ‘a mismatch between the amount of responsibility that can be attributed to any person or collection of persons on leading accounts of moral responsibility and the amount that robust intuitions suggest should be allocated to someone in a case’.⁴⁶⁶ Matthias at times appears to endorse this kind of interpretation of the alleged problem when he writes that ‘there is an increasing class of machine actions, where the traditional ways of responsibility ascription are not compatible with our sense of justice and the moral framework of society’.⁴⁶⁷ According to Köhler and others, indeed, ‘responsibility gaps’ cases purport to show that ‘our ordinary conceptual scheme for assigning responsibility, including degrees of responsibility, is seriously inadequate’ which ‘would clearly force us to seriously rethink our everyday conception of responsibility’.⁴⁶⁸ Do cases like Massacre result in ‘responsibility gaps’ of this kind? Do such cases, in other words, represent counterexamples to our best theories of responsibility so as to prompt their revision?

To address this question, we first need to get clear on what we mean by ‘responsibility’.

The term is notoriously polysemous.⁴⁶⁹ Consider the following example:

⁴⁶⁴ Miller (2007: 85).

⁴⁶⁵ Da Silva (2024); Matthias (2004); Köhler *et al.* (2018); Tigard (2020).

⁴⁶⁶ Da Silva (2024: 1).

⁴⁶⁷ Matthias (2004: 177).

⁴⁶⁸ Köhler *et al.* (2018: 55).

⁴⁶⁹ See, e.g., Bovens (1998); Collingridge (1980); Feinberg (1970); Hart (1968); Kaiserman (2021b: 3599-3603); Pesch (2015); van de Poel & Sand (2018); Zheng (2019). For an alternative way of distinguishing ‘responsibility gaps’: Santoni de Sio and Mecacci (2021). C.f. Dignum (2020).

‘A short-circuit was responsible for a signal failure on the Underground; so although I was responsible for making sure Julie got to school on time, I’m not responsible for Julie’s failure to get to school on time.’⁴⁷⁰

This toy sentence contains three occurrences of ‘responsible’. The first occurrence is an instance of ‘responsibility’ describing a purely causal relationship. The second occurrence denotes role-based duties which accrue, for example, to Julie’s parents or carers. This dissertation will not examine these first two meanings of ‘responsibility’ in any depth. The third occurrence, by contrast, picks out ‘a certain kind of relation between agents and token actions, omissions, or outcomes’,⁴⁷¹ which is often called ‘moral responsibility’. There are at least two distinct concepts which ‘responsibility’ in this third sense might pick out. On one hand, it might mean that it is not apt for me to be asked to justify, explain or generally ‘answer for’ my role in Julie’s failure to get to school on time. Let us call this concept ‘answerability’. On the other hand, it might mean that it is not apt to blame, punish or sanction me for her failure to get to school on time. Let us call this concept ‘blameworthiness’. This chapter takes each concept of moral responsibility in turn. I show that there are no discrepancies between our best theories of answerability or blameworthiness and our intuitions; illusions to the contrary trade on equivocation between the different senses of moral responsibility.⁴⁷²

7.1. Moral responsibility as answerability

Many philosophers characterise moral responsibility in terms of answerability. For Smith, ‘a morally responsible agent is one who can intelligibly be asked to ‘answer for’ her attitudes and

⁴⁷⁰ Kaiserman (2021b: 3599).

⁴⁷¹ *Ibid.* This is a sense of responsibility which some authors have referred to as ‘backwards-looking’ (Young 2011; Zheng 2019). I consider some alternative, more ‘forwards-looking’ conceptions of responsibility in the next chapter.

⁴⁷² This is the first sustained defence of two of our best theories of moral responsibility against allegations of ‘gaps’. Köhler *et al.* (2018) argue against ‘gaps’ in ‘accountability’, where accountability denotes the fittingness of some kind of sanctioning behaviour (*ibid.*: 52; Shoemaker 2011, 2015). Tigard (2020) critiques that narrow focus but addresses different theories of moral responsibility only to show that ‘gaps’ are not unique to cases involving LAWs. Kiener (2025) argues against ‘gaps’ in a theory of strict moral answerability.

conduct’⁴⁷³ where to ‘answer for’ one’s attitudes or conduct is ‘to give her (justificatory) reasons for thinking, feeling, or acting in the way she has’.⁴⁷⁴ In a similar vein, Scanlon states that someone is morally responsible for his conduct when ‘it is appropriate in a general sense to ask him to defend it or to disown it’⁴⁷⁵ and Shoemaker believes that at least one important type of moral responsibility ‘is something agents have in virtue of their ability (in principle) to ‘answer for’ their actions.’⁴⁷⁶ Recently, Kiener has articulated a novel account of this kind of moral responsibility, which he calls ‘strict moral answerability’, according to which cases like Massacre have no tendency to generate ‘gaps’ (7.1.1).⁴⁷⁷ I critically examine his theory and argue for an alternative, more robust analysis of answerability into three conditions: a causal one, an epistemic one and a control one (7.1.2). In the next section, I show that contrary to popular opinion, even on that more robust analysis of answerability which I defend, the theory contains no ‘gaps’.

7.1.1. No gaps, but abundance

In his 2024 paper, Kiener introduces a novel analysis of moral responsibility as strict moral answerability in response to what he calls the ‘Lorry Driver Paradox’. This starts from considering Williams’ lorry driver thought experiment involving a ‘lorry driver who, through no fault of his, runs over a child.’⁴⁷⁸ Williams points out that, unlike equally faultless bystanders, the driver is likely to feel ‘agent-regret’,⁴⁷⁹ that is, regret ‘which a person can feel only towards his own past actions (or, at most, actions in which he regards himself as a participant)’.⁴⁸⁰ Although ‘rightly, people will try, in comforting him, to move the driver from

⁴⁷³ Smith (2015: 103).

⁴⁷⁴ *Ibid.*

⁴⁷⁵ Scanlon (2000: 289).

⁴⁷⁶ Shoemaker (2015: 64).

⁴⁷⁷ Kiener (2024, 2025).

⁴⁷⁸ Williams (1981: 28).

⁴⁷⁹ *Ibid.*: 27.

⁴⁸⁰ *Ibid.*

this state of feeling’⁴⁸¹ to one more akin to that of a spectator, agent-regret seems morally appropriate: ‘indeed some doubt would be felt about a driver who too blandly or readily moved to that position [of a mere spectator]’.⁴⁸² At a minimum, it is usually agreed that the driver ought to ‘offer the bereaved family not only his condolences but also his deepest apology’.⁴⁸³ Kiener contends that one ought to give this kind of apology only if one is morally responsible for the outcome.⁴⁸⁴ The alleged paradox arises because, on Kiener’s view of the control and epistemic conditions for moral responsibility, the lorry driver is not morally responsible for the outcome: ‘averting the accident [was] impossible [and] he could not possibly have foreseen this accident’.⁴⁸⁵ So, Kiener contends, we need a new theory of moral responsibility:

‘the Lorry Driver Paradox [is] a strong challenge to our thinking about moral responsibility [which should prompt us] to develop a solution based on the idea of strict moral answerability, which means, roughly, answerability without any requirement of fault’.⁴⁸⁶

The new theory of moral responsibility which Kiener proposes starts from the standard conception that the kind of moral responsibility at issue, namely answerability, is ‘a request for justification, for an account of someone’s motivating reasons, aims, and values, or simply for an explanation of what good they envisaged as arising from their conduct.’⁴⁸⁷ Whilst endorsing this standard conception, Kiener proposes a novel sufficient condition:⁴⁸⁸

⁴⁸¹ *Ibid.*

⁴⁸² *Ibid.*: 28.

⁴⁸³ Sliwa (2019: 67-68); c.f. Wolf (2001: 13).

⁴⁸⁴ Kiener (2024: 363-365).

⁴⁸⁵ *Ibid.*: 362. See *infra* for criticism of Kiener’s characterisation of the control and epistemic conditions for moral responsibility.

⁴⁸⁶ *Ibid.*: 361. For an earlier account of answerability: Duff (2009). For criticism of Duff’s account: Husak (2010).

⁴⁸⁷ Kiener (2023: 204).

⁴⁸⁸ There are important differences between Duff’s and Kiener’s proposals, especially concerning the role of uncertainty in the theory and the scope of strict moral answerability, including the interpretation of Thomson’s Harm Thesis. But it is besides my purposes to consider such subtle nuances here. For a full discussion: Kiener (2024: 375-382).

‘a person is morally answerable for harm if (i) the person caused [or causally contributed to] the harm (causation clause) and (ii) (s)he had a prospective obligation to guard against such harm (obligation clause).’⁴⁸⁹

Note: the account presupposes that the person under consideration is adequately sensitive to reasons for and against the conduct in question although there is no requirement that they acted in any way negligently, recklessly or intentionally.⁴⁹⁰ Kiener applies this account to Williams’ lorry driver thought experiment, concluding that the lorry driver is morally responsible, in the answerability sense, for the outcome:

‘He caused harm through his own driving (causation clause), and, qua driver, he had a prospective obligation to guard against that very harm, that is, [the] harm of running over people with his vehicle (obligation clause). The driver also possessed the ability to be guided by the relevant reasons.’⁴⁹¹

Since *a fortiori* he was not in any way at fault, the driver is not blameworthy. However, Kiener believes that a duty to apologise only depends upon answerability.⁴⁹² So, Kiener concludes, the alleged paradox is resolved.

In his 2025 paper, Kiener applies the theory of moral responsibility as strict moral answerability to alleged ‘responsibility gaps’ cases, that is, ‘situations in which no one is morally responsible [on a standard theory] for the harm caused by AI, because no one seems to have the required control or foresight to be morally responsible for the harmful outcomes of AI systems’.⁴⁹³ For example, he reflects on ‘a fatal accident with an autonomous vehicle, [where] many people across different stages of the vehicle’s development and use played a role.’⁴⁹⁴ On

⁴⁸⁹ *Ibid*: 367. Causal contribution is addressed in Kiener (2025: 362-367).

⁴⁹⁰ Kiener (2024: 367).

⁴⁹¹ *Ibid*: 368.

⁴⁹² *Ibid*: 367.

⁴⁹³ Kiener (2025: 357).

⁴⁹⁴ *Ibid*: 362.

his account of moral responsibility, Kiener argues that such cases do not evince ‘gaps’ in the theory of moral responsibility but rather ‘*abundance*, that is, a situation in which numerous agents are responsible for the harm caused by AI’.⁴⁹⁵ This is because of the large number of individuals who contribute to the causation of that harm and who had ‘specific obligations, tailored to their position and expertise’ to guard against such harms.⁴⁹⁶ There are several problems with Kiener’s account of moral responsibility, which prompt me to amend his theory in 7.1.2; but my findings will support that conclusion that answerability contains no ‘gaps’.

7.1.2. Epistemic and control conditions

The causation clause which Kiener defends is intuitively plausible and relatively uncontroversial: it is commonly assumed that agents bear responsibility only for those outcomes which they ‘bring about’,⁴⁹⁷ that is, ‘there must have been some sort of causal connection between [the agent’s] action or omission and the harm’.⁴⁹⁸ My principal objection is to the obligation clause.

It is unclear what justificatory work the obligation clause is capable of doing in the ascription of moral responsibility without circularity. Kiener introduces the obligation clause to demarcate those who merely causally contribute to an outcome’s production from those who are morally responsible, in the answerability sense, for the outcome.⁴⁹⁹ However, Kiener’s appeal to the idea of what one has a prospective obligation to guard against seems to rely upon the idea that such obligations are attached to jobs or social roles:

⁴⁹⁵ *Ibid*: 357-358.

⁴⁹⁶ *Ibid*: 368.

⁴⁹⁷ Fischer and Ravizza (1998: 99-100, 106, 134).

⁴⁹⁸ Feinberg (1970: 195).

⁴⁹⁹ See, e.g., Kiener’s (2024) responses to problematic examples of accidentally knocking over a vase hidden behind a door or unwittingly causing someone to be electrocuted by flicking a light switch.

‘everyone in the development and use of AI shares the general obligation to guard against certain risks of harm and [...] they possess some more specific obligations, tailored to their position and expertise’.⁵⁰⁰

In my submission, this appeal to the specific obligations of persons within the AI industry cannot justify a distinction between purely causal responsibility and moral responsibility based on the obligation clause since Kiener is interested in precisely those specific obligations which have not been breached:

‘whenever [an] AI system causes harm, many of those involved in the development and use of this AI system are likely to become strictly morally answerable for the harmful outcome’.⁵⁰¹

In his defence of the obligation clause, Kiener appeals to Gardner’s view that when we fail to conform to reasons which apply to us, those reasons persist in exerting a certain normative force.⁵⁰² But again, this is a *non sequitur*: *a fortiori* those involved in the development and use of AI systems in the faultless accidents that Kiener is interested in *have not failed* to conform to the reasons which apply to them; quite the contrary, they have fully discharged both their general obligation to ‘guard against certain risks of harm’ and their ‘more specific obligations, tailored to their position and expertise’.⁵⁰³ How, then, is the obligation clause to be motivated? And if the obligation in question has been entirely discharged, how can it generate a moral duty to answer for one’s conduct which is borne only by a subclass of the causal contributors, namely those whose job it is to guard against such harms? How can we justify allocating moral

⁵⁰⁰ Kiener (2025: 368).

⁵⁰¹ *Ibid*: 369.

⁵⁰² Gardner (2005).

⁵⁰³ Kiener (2025: 368).

responsibility to some rather than others if we argue in a circle that some rather than others are morally responsible because that is their social role?

One option is for Kiener to abandon the obligation clause and embrace an account of answerability as purely causal responsibility. But this would generate some surprising consequences. One reason is that there is usually taken to be an epistemic condition on moral responsibility which captures the intuition that moral responsibility does not extend to outcomes which the agent ‘couldn’t reasonably have foreseen’.⁵⁰⁴ For instance, if P ‘take[s] her dog for a walk around the park on a bright day, and the dog is later struck by lightning’,⁵⁰⁵ or ‘toss[es] an orange pip over the edge of a cliff and in doing so precipitate[s] a rock fall’,⁵⁰⁶ then P’s mere causal connection to the outcome begets no moral responsibility.⁵⁰⁷

So: I suggest that we replace Kiener’s obligation clause by an epistemic condition distinguishing those for whom an outcome like Massacre was in some sense ‘reasonably foreseeable’⁵⁰⁸ from those who, though causally connected by their actions or omissions to its production, could not reasonably have foreseen that kind of result. Consider, by way of illustration:

Cancer Researcher: It turns out that the code which went wrong in Massacre derived from Cancer Researcher’s seemingly unrelated work concerning the automation of cancer research.

⁵⁰⁴ Kaiserman (2021a: 144); c.f. Miller (2004: 245). For discussion of the epistemic condition: Feinberg (1986: 269-315); Robichaud and Wieland (2017).

⁵⁰⁵ Kaiserman (2021b: 3603).

⁵⁰⁶ Miller (2004: 245).

⁵⁰⁷ Kiener assumes that the outcome in the lorry driver case is one which the driver could not reasonably have foreseen. I challenge this assumption in 7.3.

⁵⁰⁸ See *infra* 7.2.1.

Hunter-Killer Drones Researcher: It turns out the code which went wrong in Massacre came from research published in the LAWs literature by Hunter-Killer Drones Researcher.

It seems that Hunter-Killer Drones Researcher may appropriately be called to answer morally for her role in the outcome's production in Massacre. But the intuition is less clear as regards Cancer Researcher. I suggest that this is because it was a reasonably foreseeable consequence of Hunter-Killer Drones Researcher's work that her research might be causally connected to eventual harm of this kind. But the role which Cancer Researcher's work plays in Massacre seems more like that of the tossed orange pip precipitating a rockfall. Note: it does not follow that either Cancer Researcher or Hunter-Killer Drones Researcher is at *fault* or *blameworthy* for what happened, provided that each took every reasonable precaution before publishing their code. So, the account of answerability remains, in that sense, 'strict'.

My final objection to Kiener's account of answerability is in part a question of truth-in-advertising. In essence, Kiener makes the account sound 'stricter' than it is by misrepresenting the standard control condition for moral responsibility. Indeed, Kiener's case for a new theory of moral responsibility as strict moral answerability assumes that Williams' lorry driver 'does not satisfy a control condition'⁵⁰⁹ for moral responsibility on standard theories because 'averting the accident [was] impossible'.⁵¹⁰ But the assumption that an agent must control whether or not the outcome occurs in order to bear moral responsibility for it is, as I develop in the next section, false. On the contrary, Frankfurt-style cases⁵¹¹ and the moral luck literature⁵¹² foreground the possibility of freely deciding to do something in a way which seems sufficient for moral responsibility even if one has no control over whether the outcome occurs

⁵⁰⁹ Kiener (2024: 362).

⁵¹⁰ *Ibid.*

⁵¹¹ Frankfurt (1969). For critical engagement: Sartorio (2016).

⁵¹² Nagel (2012: 25): there is inescapable 'luck, good and bad, in the way things turn out'.

due to backup causes which ensure that, respectively, my decision to act or the outcome itself could not have occurred otherwise. As such, modern accounts often ground the control requisite for moral responsibility in facts about whether and to what extent the agent was responsive, at the point of causally contributing, to reasons for or against acting or omitting to act. For example, Fischer and Ravizza argue that agent P satisfies the control condition only if there is some possible world at which P acts, via the same mechanism that actually produced her action, in accordance with the reasons that she has.⁵¹³ Williams' lorry driver satisfies this kind of control condition to the extent that he was responsive in deciding to drive to reasons for and against driving. My objection then is that Kiener 'strawmans' the standard theory of moral responsibility by assuming that it requires control over outcomes. Then, he presupposes a standard control condition as part of his account of strict moral answerability, namely that 'the person under consideration possesses the ability to be guided by the relevant reasons, [viz] the ability to recognize and respond to the reasons both in favor of and against the conduct under consideration.'⁵¹⁴ This 'control' condition is—in addition to the causal and epistemic conditions which I have glossed—a third necessary condition for the appropriate attribution of answerability.

I have critically examined Kiener's argument for strict moral answerability and argued for the reinstatement of standard epistemic and control conditions in the theory of answerability. However, Kiener presumes that any theory of moral responsibility founded upon standard control and epistemic conditions contains 'gaps' in cases like Massacre.⁵¹⁵ This prompts the question: do we still get an 'abundance' of answerability in cases like Massacre on

⁵¹³ For criticism: McKenna (2005); Watson (2001). For an alternative 'agent-sensitive' approach: Sartorio (2016). For a causal account of the idea that reasons-sensitivity can come in degrees: Kaiserman (2020).

⁵¹⁴ Kiener (2024: 367).

⁵¹⁵ Kiener (2025: 357).

the more robust account of answerability which I have defended? In the next section, I show that we do.

7.2. Look: no gaps!

In the last section, I sketched an analysis of moral responsibility as answerability according to which an agent is answerable for all and only those outcomes to which she is appropriately causally and epistemically connected by her reasons-responsive actions or omissions. In this section, I show that the theory of answerability which I have adumbrated gets the right answers in cases like Massacre. Neither the unpredictability (7.2.1) nor the uncontrollability (7.2.2) of LAWs has any tendency to undermine the answerability of those who produce, sell and use LAWs in cases like Massacre. Nor are there any ‘breaks’ in the causal chain leading from the actions or omissions of those who produce, sell and use LAWs to the outcome in such cases (7.2.3).

7.2.1. The epistemic condition

A number of authors have suggested that our best theories of moral responsibility get the wrong answers in cases like Massacre, generating ‘gaps’, at least in part because those who we intuitively think are morally responsible for the outcome do not satisfy the standard epistemic conditions for moral responsibility.⁵¹⁶ For instance, it has been argued that since ‘the possibility that an autonomous system will make choices other than those predicted and encouraged by its programmers is inherent in the claim that it is autonomous [...] it will no longer be possible [to] hold the programmers/designers responsible for outcomes’⁵¹⁷ and that the moral responsibility of military commanders for the use of LAWs is ‘mitigated [...] due to epistemic constraints’;⁵¹⁸ indeed, no one should be ‘expected to ‘foresee the unforeseeable’ if the system

⁵¹⁶ Kiener (2025: 357-358); Danaher (2016a); ICRC (2014: 8, 16, 22-23); Roff (2013); Sparrow (2007). For criticism: Köhler *et al.* (2017).

⁵¹⁷ Sparrow (2007: 70); c.f. Danaher (2016a: 305).

⁵¹⁸ Roff (2013: 358).

is used within the confines provided'.⁵¹⁹ I argue that whilst LAWs are unpredictable that in no way undermines the epistemic connection between those who contribute by their actions to the production, sale and use of LAWs and the eventual outcomes.

Let us grant that LAWs like the one used in Massacre are unpredictable in the following sense. Recall that one difference between LAWs and automated weapon systems, as explained in 6.2.2-6.2.3, is that to display the requisite discretion and adaptability in a vast array of different situations, LAWs are expected to involve either unsupervised or weakly supervised deep learning. One upshot of this, just as we saw regarding SARs in Part One, is that LAWs may behave in ways which appear to be impossible for the manufacturer, seller or commander of a LAW to specifically predict.⁵²⁰ As a result, even a tried-and-tested LAW, which reliably distinguishes between military targets and civilian objects, may behave unexpectedly in any new situation.⁵²¹ But to infer that those who freely and knowingly contributed to the production, sale and use of the LAW are not answerable for the outcome in cases like Massacre depends, in my submission, upon at least three mistakes concerning the nature of epistemic connection required for answerability.⁵²²

First, the epistemic condition for answerability is not a purely subjective matter.⁵²³ On the contrary, as mentioned earlier, it is usually stated in the philosophy literature in terms of 'reasonable foreseeability' rather than actual foresight.⁵²⁴ The term 'reasonable foreseeability'

⁵¹⁹ ICRC (2014: 46).

⁵²⁰ Matthias (2004: 177); Sparrow (2007: 70-71).

⁵²¹ *Ibid.*

⁵²² Traditionally, the epistemic condition on responsibility was presumed to be the 'relatively straightforward and unproblematic' condition (Rudy-Hiller 2025). Strictly, there are at least four interwoven kinds of epistemic conditions which Rudy-Hiller (2025) calls requirements of 'awareness'. The one which is relevant for our purposes is 'awareness of consequences'.

⁵²³ The classic example in the philosophy literature is that of a doctor who fails to attend mandatory training and thereby fails to actually foresee a reasonably foreseeable harm to a patient by the administration of a drug. See, e.g., Kaiserman (2021a: 144); Wieland (2017). In such cases, it would seem bizarre not to call the doctor to answer for, explain or justify her conduct on the basis that she had not actually foreseen the injury.

⁵²⁴ Vargas (2005); Fischer and Tognazzini (2009); Sartorio (2017); c.f. Zimmerman (1986); Nelkin and Rickless (2017). Note that the philosophy literature cited often does not distinguish between answerability and

derives from the case law in the law of torts, which contains extensive discussion of the appropriate interpretation of that term.⁵²⁵ A classic statement, albeit one which fails to analyse ‘reasonable’ itself, derives from the ‘neighbour principle’ in *Donoghue v Stevenson* according to which reasonably foreseeable harms resulting from my actions or omissions are those which injure:

‘persons who are so closely and directly affected by my act that I ought reasonably to have them in contemplation as being so affected when I am directing my mind to the acts or omissions which are called in question.’⁵²⁶

Of course, the case law on the interpretation of a particular term should not be decisive in a philosophy thesis, but it may be persuasive. Indeed, I suggest that non-combatants are ‘so closely and directly affected’ by the decisions of all those who produce, sell and use weapons systems that the latter ‘ought reasonably to have them in contemplation as being so affected’ when acting or omitting to act in the relevant ways.⁵²⁷ So harms to non-combatants in cases like Massacre arguably fall within the scope of a common law interpretation of reasonable foreseeability. Importantly, it does not follow that anyone in Massacre was negligent. The principle stated in *Donoghue v Stevenson*, and developed in subsequent case law,⁵²⁸ is merely to take ‘reasonable care’ with respect to mitigating reasonably foreseeable harms.⁵²⁹ *A fortiori* reasonable care with respect to mitigating reasonably foreseeable harms like those in Massacre

blameworthiness. Because I view answerability as a necessary but not sufficient condition for blameworthiness (Duff 2009), this is not a problem for the view of answerability which I defend.

⁵²⁵ For a history of the concept of foreseeability examining *Donoghue v Stevenson* [1932] AC 562 at 580 and *Robinson v Chief Constable of West Yorkshire Police* [2018] AC 736: Scott (2019).

⁵²⁶ *Donoghue v Stevenson* at 580; c.f. *Heaven v Pender* (1883) 11 QBD 503 at 509.

⁵²⁷ This may explain the origin of some of those general and specific obligations to ensure that weapon systems operate safely and in accordance with international legal rules which attach to jobs in the production, sale and use of LAWs.

⁵²⁸ See *Caparo Industries Plc v Dickman* [1990] 2 AC 605; c.f. Lord Reed’s criticisms of the *Caparo* test for a duty of care in *Robinson v Chief Constable of West Yorkshire Police* at [21]-[30].

⁵²⁹ *Donoghue v Stevenson* at 580.

has been taken. But that, as I explain in 7.3, goes to the point that none of those who are *answerable* for the outcome in Massacre appear to be *blameworthy*.

One might object that the harms in Massacre were not reasonably foreseeable because no one who contributed to the production, sale and use of the LAW could possibly have predicted how events in Massacre would unfold. The AI went rogue! Indeed, there are philosophers who believe that for an outcome to be a reasonably foreseeable consequence of an agent's action or omission, it must be the case that an event of *exactly* that kind was a reasonably foreseeable consequence.⁵³⁰ However, others believe that the scope of reasonable foreseeability is more coarse-grained.⁵³¹ Once again, the position at common law will not be decisive but it may be persuasive of the merits of the coarse-grainers. At common law, an outcome is reasonably foreseeable if it is of a type or kind which was reasonably foreseeable in the circumstances, even if it occurred in a manner which was not reasonably foreseeable.⁵³² For example, in *Jolley v Sutton London Borough Council*, a 14-year-old boy suffered severe spinal injuries, resulting in paraplegia, when a rotten boat abandoned on council land fell on him whilst he was attempting to fix it. The House of Lords held that it was a reasonably foreseeable consequence of the council's failure to remove the boat that children would 'meddle with the boat at the risk of some physical injury'.⁵³³ Although the way in which events unfolded was unforeseeable, the injury fell within that general kind of outcome which was reasonably foreseeable in the circumstances.⁵³⁴ I suggest that a similar idea should apply in Massacre. Although the specific outcome, including the particular way in which events unfolded, was not foreseeable at the time when the relevant parties contributed to the

⁵³⁰ Zimmerman (1986); Vargas (2005).

⁵³¹ Fischer and Tognazzini (2009); King (2017: 272); Nelkin and Rickless (2017: 126).

⁵³² *Jolley v Sutton London Borough Council* [2000] 1 WLR 1082 at 1090; *Rushbond v JS Design* [2021] EWCA Civ 1889 at [64]. C.f., *The Wagon Mound (No.1)* [1961] AC 388.

⁵³³ *Jolley v Sutton London Borough Council* at 1093.

⁵³⁴ *Ibid.*

production, sale and use of the LAW, it *was* reasonably foreseeable that an outcome of that general kind would occur, particularly given what those parties reasonably know about that potential for unpredictable accident which is intrinsic to machine learning systems of that kind.

One might at this juncture counter that an outcome of that general kind was not reasonably foreseeable because *a fortiori* the LAW is tried-and-tested and operates reliably in a safe way. Even this general kind of outcome was entirely unforeseeable because it was so improbable. So goes at least one kind of objection in the literature.⁵³⁵ But consider:

Russian Roulette: P, Q and R respectively produce, sell and use a Russian Roulette machine for a magic show which is triggered by a random number generator, generating numbers between 0 and 1 billion. The machine will shoot V if and only if the number generated is 7. The machine is tried-and-tested and operates reliably in a safe way in the practice sessions. On stage, the random number generator lands on a 7 and V dies.

Intuitively, V's death was a reasonably foreseeable consequence of this type of magic show.⁵³⁶ Nonetheless, the machine was tried-and-tested, operated reliably in a safe way in practice sessions and, by all accounts, the outcome was very unlikely. I do not mean to suggest that the production, sale and use of LAWs is a gratuitous game of risk like Russian Roulette. But the example serves to highlight that very improbable, harmful outcomes are within the reasonable foresight of those who produce, sell and use machines which carry such an intrinsic risk. It bears repeating: this does not mean that the production, sale and use of LAWs is in breach of any duty of care, although that might be the case in Russian Roulette. As such, those who

⁵³⁵ ICRC (2014: 14).

⁵³⁶ In the case law on negligence, probability is sometimes referred to in determining whether a risk was reasonably foreseeable. For example, in *Dean & Chapter Of Rochester Cathedral v Leonard Debell* [2016] EWCA Civ 1094, counsel described the accident as a 'fantastic possibility' (that anyone would walk so close to the bollard and trip). In finding that the cathedral should not be liable for the accident, Elias LJ held that the concept of reasonable foreseeability involves a 'practical and realistic approach' to the kinds of dangers which the cathedral were obliged to remedy. I suggest that in fact the issue of whether the accident was reasonably foreseeable (arguably, it was) was conflated in that case with the issue of whether the cathedral had been negligent (it had not). For discussion: Omar (2017); c.f. Bagshaw (2024).

produce, sell and use the LAW are not *blameworthy* for their part in Massacre. They are, however, answerable.⁵³⁷

7.2.2. The control condition

Another way in which authors have sought to argue for ‘gaps’ in our best theories of moral responsibility is to claim that whilst intuitively some agent is morally responsible for the outcome in cases like Massacre, that agent does not satisfy the standard control conditions for moral responsibility.⁵³⁸ For example, Sparrow fears that those who produce, sell and use LAWs may be unfairly held responsible ‘for the actions of machines whose decisions they did not control’.⁵³⁹ Santoni de Sio and van den Hoven suggest that ‘human responsibility [would] evaporate’ on standard accounts of the control condition for responsibility ‘if systems are acting on their own accord, [with] little human control and intervention’.⁵⁴⁰ On this basis, Santoni de Sio and van den Hoven articulate and defend an alternative conception of the control condition for responsibility, which they call ‘meaningful human control’, and indicate that LAWs should be designed to ensure the satisfaction of that condition. This kind of approach aims rather to prevent the alleged conceptual problem for our theories of responsibility from raising its head in practice than to critically engage with it.⁵⁴¹ By contrast, Himmelreich articulates and defends an alternative account of the control condition for responsibility, arguing that ‘there is a plausible sense of ‘control’ on which it is true that the commander has control over something related to the bombing’ in cases like Massacre.⁵⁴² On all accounts, however, the underlying assumption appears to be that, on standard accounts of the control condition, an agent is not

⁵³⁷ It might be objected that what people mean in arguing for ‘gaps’ in our theory of responsibility must therefore be blameworthiness, not answerability. But I show in 7.3 that there are no gaps in blameworthiness either. The apparent problem seems to trade on equivocation between blameworthiness and answerability.

⁵³⁸ Kiener (2025: 357, 2022b); c.f. Coeckelbergh (2020); Nyholm (2020).

⁵³⁹ Sparrow (2007: 70-71).

⁵⁴⁰ Santoni de Sio and van den Hoven (2018: 1-2).

⁵⁴¹ C.f. Köhler *et al.* (2017); Eggert (2024).

⁵⁴² Himmelreich (2019: 736).

responsible for an outcome if she did not exercise the appropriate kind of control over some part of the causal sequence leading to its production.⁵⁴³ I will show that this assumption is false; it is sufficient for answerability that my actions or omissions in causally contributing to the production of some reasonably foreseeable outcome were themselves reasons-responsive.

Recall Russian Roulette. In that case, P, Q and R each seem to be answerable for V's death but none of them exercises control over the Russian Roulette machine. I suggest that what motivates the intuition that each of P, Q and R is answerable is simply that at the time of their causal contributions to the production, sale and use of the machine, each of them was reasons-responsive, namely they 'possesse[d] the ability to be guided by the relevant reasons, [viz] the ability to recognize and respond to the reasons both in favor of and against the conduct under consideration.'⁵⁴⁴ Contrast:

Drugged Russian Roulette: The case is identical to Russian Roulette but in this case P, Q and R have been tricked into consuming drugs which make them contribute to the production, sale and use of the Russian Roulette machine.

In that case, I suggest that P, Q and R are *not* reasons-responsive at the point of their causal contribution to the production of V's death. As such, they are not in control of their actions or omissions at the relevant time and therefore not answerable for the outcome. By contrast:

Voluntary Drugged Russian Roulette: The case is identical to Drugged Russian Roulette but in this case P, Q and R freely and knowingly took drugs which make them contribute to the production, sale and use of the Russian Roulette machine.

In Voluntary Drugged Russian Roulette, I suggest that P, Q and R *are* reasons-responsive at the point of their causal contribution to the production of V's death. This is because they are in

⁵⁴³ See, e.g., Santoni de Sio and van den Hoven (2018: 7, 9); Himmelreich (2019: 736-737).

⁵⁴⁴ Kiener (2025: 357).

control of their actions or omissions at the relevant time of taking the drugs and the outcome is reasonably foreseeable at that earlier time; they are therefore answerable for the outcome. It suffices, I submit, for control—as with the question of epistemic connection—that at the point of P’s causal contribution to the production of the outcome P was in control of those actions which causally contributed to the outcome’s occurrence and that kind of outcome was a reasonably foreseeable consequence of P’s causal contribution.

I do not wish to suggest that this is a novel point in the philosophy or legal literature. On the contrary, the idea that ‘[an] agent’s responsibility may be grounded in the fact that their failure to fulfill certain conditions on responsibility is traceable to earlier actions undertaken by the agent when they did fulfill these conditions’⁵⁴⁵ is familiar from the philosophy literature on ‘tracing’.⁵⁴⁶ Sartorio has defended the related ‘Principle of Derivative Responsibility’ according to which ‘If an agent is responsible for X, X causes Y, and the relevant epistemic conditions for responsibility obtain, then the agent is also responsible for Y’.⁵⁴⁷ This seems consonant with the idea articulated in the classic case of *DPP v Majewski*,⁵⁴⁸ where Lord Elwyn-Jones LC famously held that voluntary intoxication by drugs or alcohol is no defence, at least to ‘basic intent’ crimes.⁵⁴⁹ But it is for this very reason that I find it so bizarre in the AI ethics literature to be confronted by such a widespread assumption that a necessary condition for responsibility is that an agent ‘has control over whether an outcome [...] occurs’⁵⁵⁰ or that ‘an agent can legitimately be seen as in control of a certain outcome’.⁵⁵¹ This foundational mistake about the control condition is, I suggest, why neither Himmelreich (with his ‘robust

⁵⁴⁵ Talbert (2024).

⁵⁴⁶ Fischer and Ravizza (1998: 49-51); Khoury (2012); King (2014); Shabo (2015). The tracing strategy has been subject to criticism (Vargas 2005) and defence (Fischer and Tognazzini 2009).

⁵⁴⁷ Sartorio (2016: 76).

⁵⁴⁸ [1976] UKHL 2 at 475; c.f., *R v Lipman* 1 QB 152.

⁵⁴⁹ There is some subtlety here regarding the distinction between ‘basic intent’ and ‘specific intent’ crimes, where *mens rea* for the latter may not be established in cases of intoxication (*R v Heard* [2007] EWCA Crim 125).

⁵⁵⁰ Himmelreich (2019: 736).

⁵⁵¹ Santoni de Sio and van den Hoven (2018: 10).

tracking control’ condition⁵⁵² over probabilistic outcomes) nor Santoni de Sio and van den Hoven (with their ‘meaningful human control’ condition⁵⁵³) have articulated a necessary condition for the answerability of any party in cases like Massacre. In other words, arguments for the existence of ‘gaps’ in our best theories of moral responsibility on the basis of some kind of loss of control over outcomes involve a *non sequitur*: our best theories of moral responsibility require only reasons-responsiveness at the point of causal contribution and that kind of control, in lethal accidents involving AI, is not disputed.

7.2.3. The causal condition

Some authors have sought to argue that cases like Massacre reveal ‘gaps’ in our best theories of responsibility on the basis that the autonomous nature of the LAW in some way undermines the causal connection between the actions of those who contribute to its production, sale and use and the eventual consequences of the LAW.⁵⁵⁴ For instance, Sparrow writes that ‘[t]he connection between the programmers/designers and the results of the system, which would ground the attribution of responsibility, is broken by the autonomy of the system.’⁵⁵⁵ Danaher similarly writes that ‘a high degree of autonomous power [...] threatens to break the link between the human creators and designers of such systems and their ultimate causal effects.’⁵⁵⁶ Turner focuses on the ‘unpredictability of AI’ which he believes ‘renders it ever more difficult

⁵⁵² Additional problems with Himmelreich’s account abound. First, he articulates an account of control involving ‘regulative control’ rather than ‘guidance control’ in Fischer and Ravizza’s terminology (Himmelreich 2019: 737, n.22), which arguably falls foul of Frankfurt-style cases involving backup causes. Moreover, the sufficient condition for robust tracking control relies upon counterfactuals which are known to be vulnerable to backtracking readings (Edgington 2011: 235), erroneous ‘common cause’ readings (Kment 2010) and issues related to ‘future similarity’ of worlds (Fine 1975).

⁵⁵³ Another problem is that the authors seem to assume that reasons-responsiveness on Fischer and Ravizza’s account would not extend to cases involving LAWs because the ‘decisions’ of LAWs may not themselves be reasons-responsive (Santoni de Sio and van den Hoven 2018: 6). But this would be dealt with under Fischer and Ravizza’s concept of ‘taking responsibility’ for one’s decisional mechanism (1998: 207-239; Fischer 2004: 146) such that, as Santoni de Sio and van den Hoven (2018: 5-6) acknowledge, ‘[i]f you consciously decide to settle an important matter by the toss of a coin [...] you have to accept that others will hold you accountable for the outcome of the choice (you cannot blame the system)’.

⁵⁵⁴ Sparrow (2007: 70); Danaher (2016a: 300); Turner (2019: 70).

⁵⁵⁵ Sparrow (2007: 70).

⁵⁵⁶ Danaher (2016a: 300).

to tether each decision AI takes to humans through a traditional chain of causation’.⁵⁵⁷ The idea seems to be that, whilst intuitively those who contribute to the production, sale and use of autonomous AI systems like LAWs bear some kind of responsibility for the ‘decisions’ of those systems, there is something about the autonomous nature of a LAW which ‘breaks the chain of causation’ between the actions or omissions of those persons and the outcome in cases like Massacre. I will show that even if we can make sense of talk of chain-breaking, in the context of autonomous AI or otherwise, cases like Massacre do not seem to involve any such phenomenon. So, arguments for ‘gaps’ in answerability based on the causal condition are unsound. The argument proceeds by distinguishing three kinds of phenomena which authors might mean when they talk about ‘breaking the chain of causation’.⁵⁵⁸ There is a traditional libertarian account (7.2.3.1), an epistemic or control account (7.2.3.2) and a causal threshold account (7.2.3.3). I will suggest that the former is implausible and the latter two do not support the view that cases like Massacre involve ‘gaps’ in answerability.

7.2.3.1. The traditional libertarian account

According to the traditional libertarian account of the *novus actus interveniens* doctrine which has held sway in morality and in law, ‘the free, deliberate and informed act or omission of a human being [...] negatives causal connection’.⁵⁵⁹ The account rests on a controversial account of human decision-making according to which voluntary human decisions are uncaused events which begin new causal chains such that it is not possible to cause an outcome through the ‘free, deliberate and informed’ actions or omissions of another.⁵⁶⁰ This account of human decision-making has been critiqued as ‘difficult to sustain’⁵⁶¹ and ‘bordering on the

⁵⁵⁷ Turner (2019: 70).

⁵⁵⁸ I am indebted to Radomirescu (2022) for her analysis of chain-breaking.

⁵⁵⁹ Hart and Honore (1985: 136, italics in original).

⁵⁶⁰ *Ibid.*

⁵⁶¹ Moore (2000: 868)

unintelligible’.⁵⁶² Even if we can make sense of the account for human decision-making, it is not clear that the ‘actions’ of LAWs are relevantly similar to the ‘free, deliberate and informed act[s] or omission[s] of a human being’.⁵⁶³ The ‘decisions’ of LAWs seem to lack certain relevant features which might be taken to distinguish human agency, such as consciousness and sentience. Indeed, talk of ‘decisions’ by autonomous AI has been called an ‘anthropomorphic trope granting statistical models a degree of autonomy that dangerously downplays the true role of human agency in sociotechnical systems’.⁵⁶⁴ In counterpoint, at least one author has articulated and defended a libertarian view of agency according to which autonomous AI systems might constitute agents.⁵⁶⁵ Why, then, would the ‘decisions’ of LAWs not ‘break the causal chain’ in cases like Massacre?

7.2.3.2. The epistemic or control account

The principal reason why the ‘decisions’ of LAWs would not ‘break the causal chain’ is not in my view that LAWs cannot be agents. The principal reason is that agency does not really seem to matter *per se* for chain-breaking. This is also my principal reason for taking the view that the traditional libertarian account of chain-breaking should be rejected, at least for the purposes of ascribing answerability or blameworthiness.⁵⁶⁶ Consider, by way of illustration, the following cases:

Bomb: P lights a fuse on a bomb; once lit the fuse cannot be extinguished. The fuse burns out and V is killed in the explosion.⁵⁶⁷

⁵⁶² *Ibid*: 871.

⁵⁶³ Hart and Honoré (1985: 136).

⁵⁶⁴ Watson and Floridi (2020: 9216).

⁵⁶⁵ Steward (2012: ch.4) propounds a libertarian view of agency which allows for human, animal and algorithmic agency in sufficiently complex systems.

⁵⁶⁶ This kind of view has been defended in full by Kaiserman (2021b), Moore (2000) and Zimmerman (1985); I do not purport to offer a full defence of it here.

⁵⁶⁷ This example and the following ones are from Kaiserman (2021b).

Bomb (Intervening Agent): P plants a bomb and lights the fuse but the fuse is extinguished by a gust of wind. The fuse is relit by a passing opportunistic arsonist and V is killed in the explosion.

Bomb (Lightning): P plants a bomb and lights the fuse but the fuse is extinguished by a gust of wind. The fuse is relit by a lightning strike and V is killed in the explosion.

P seems to be fully responsible for V's death in Bomb. However, P's responsibility for V's death seems to be diminished in Bomb (Intervening Agent) and in Bomb (Lightning). Some authors have suggested that this is due to the occurrence of an event which was beyond P's reasonable foresight or control in each case. For example, Zimmerman has argued that there is 'nothing special about intervening agents with respect to the allocation of [responsibility]';⁵⁶⁸ rather, 'the actions of intervening agents are [often] difficult to foresee and control, and this renders many of the [...] consequences of our own actions [in cases with intervening agents] difficult to foresee and control'.⁵⁶⁹ Similarly, Hart and Honoré note that 'an act of God, in the sense of an event whose occurrence at the place in question is extraordinary' can 'relieve a wrongdoer of responsibility' in the same way that acts of intervening agents can.⁵⁷⁰ This kind of interpretation of chain-breaking seems more promising than the traditional libertarian account.⁵⁷¹ However, it would not apply in cases like Massacre: the 'actions' of the LAW are not an act of God, nor are they beyond the reasonable foresight or control of those who contributed to its production, sale or use. Quite the contrary. As discussed in 7.2.1-7.2.2, the LAW's 'decision' was reasonably foreseeable and my failure to control some aspect of a causal

⁵⁶⁸ Zimmerman (1985: 357).

⁵⁶⁹ *Ibid.*: 356.

⁵⁷⁰ Hart and Honoré (1985: 162–3). There is some complexity here which I gloss over: Hart and Honoré are discussing a legal concept of responsibility which is somewhat ambiguous between answerability and blameworthiness, as Duff (2009) expounds on.

⁵⁷¹ This explanation fits better with the case law on chain-breaking according to which, where there is a duty to take reasonable care to protect against actions of a third party, third party actions consequent upon the failure to take such care are not a *novus actus interveniens* (*Reeves v Commissioner of Police of the Metropolis* [1999] 3 WLR 363 at 367-368; *Stansbie v Troman* [1948] 2 KB 48 at 52).

chain leading from my action or omission does not detract from my answerability for the outcome provided that I was appropriately in control of my earlier action or omission.

7.2.3.3. The causal threshold account

An alternative interpretation of cases like Bomb (Intervening Agent) and Bomb (Lightning) has been defended by Moore⁵⁷² and Kaiserman.⁵⁷³ These authors have argued that an intervening event is judged to break the chain of causation in cases where it relegates the initial agent's causal contribution to the outcome to such a small degree⁵⁷⁴ that it falls below some *de minimis* threshold. Kaiserman has argued further that we should abandon this idea of a threshold and have a smoothly varying concept of responsibility in morality as in law to reflect the full scale of causal contributions. For example, in Bomb (Intervening Agent) and Bomb (Lightning), we might conclude that P's planting the bomb *together* with the opportunistic arsonist or the lightning strike (respectively) caused V's death.⁵⁷⁵ But suppose for the time being that we keep the threshold.⁵⁷⁶ In cases like Massacre, do the actions or omissions of the producers, sellers or users of LAWs fall below some *de minimis* threshold for causal contribution?

Because this is a philosophy thesis, I will not offer any general rule for determining the threshold of significance for causal contribution to the production, sale and use of a LAW.⁵⁷⁷ In practice, the actions or omissions of *some* of those who work on the production, sale and use

⁵⁷² Moore (2000).

⁵⁷³ Kaiserman (2021b).

⁵⁷⁴ Kaiserman (2016, 2017a, 2017b), Kaiserman and Beebee (2019) and Braham and van Hees (2012) offer further analysis of this notion of degrees of causal contribution.

⁵⁷⁵ See Kaiserman (2016, 2017a, 2021a, 2021b).

⁵⁷⁶ Kiener (2025: 367) assumes a smoothly varying concept of responsibility combined somewhat strangely with a binary concept of answerability: 'degrees of responsibility do not affect that a person is answerable but may change how strong a person's obligation or reasons to answer are'. As such, 'the minuteness of a contribution does not preclude one's strict moral answerability in cases of accidents'. This seems to me an unnecessary multiplication of concepts: on a no-threshold view like Kiener's, answerability just is a function of the strength of a person's obligation or reasons to answer.

⁵⁷⁷ See Kaiserman and Beebee (2019) for a defence of this type of approach in the analysis of causal contribution in war.

of the LAW in Massacre will fall below the threshold. Indeed, some, by moral luck, will have been poor workers who gave nothing to the collective project. Those people do not bear answerability for the outcome. But not everyone will get off the hook. For one thing, on a view of agency which allows for group agency, there are a number of group agents which will have made significant causal contributions to the production, sale and use of the LAW in just the same way in which BAE systems may be said to have made significant causal contributions to the production, sale and use of weapons used to perpetrate war crimes in Yemen.⁵⁷⁸ For another thing, even if we do not allow for group agency, many causal contributions to the production, sale and use of the LAW will have been significant.⁵⁷⁹ Indeed, there are agents who made key decisions about the type of machine learning to be used, the nature and quality of the training data, the sale of the LAW and its ultimate deployment. Those agents and others who make more-than-*de-minimis* causal contributions to the production, sale and use of the LAW are answerable for the outcome, provided that they satisfy the epistemic and control conditions examined above. As Kaiserman and Beebe argue in their analysis of liability to defensive harm in war:

‘It might be true that in general those who play a particular kind of role [...] are likely to meet or fail to meet the required threshold; but from a metaphysical point of view such rough-and-ready generalisations can only serve as a best guess, in a situation of extreme and inevitable epistemic uncertainty [...].’⁵⁸⁰

⁵⁷⁸ Merat (2019).

⁵⁷⁹ It might be objected that it is in principle possible that every person who causally contributed to the production, sale and use of the LAW made a *de minimis* contribution which fell below the threshold. To the extent that this is a feature of the analysis of chain-breaking proposed here, it seems to be more a reason to reject the idea of there being a threshold for the significance of causal contribution than to treat cases like Massacre as counterexamples to our best theories of responsibility.

⁵⁸⁰ Kaiserman and Beebe (2019: 375). As to the responsibility of some civilians for cases like Massacre: Roff (2013: 358).

So, I suggest that whether we understand chain-breaking in terms of reasonable foreseeability, control or a causal threshold, there is likely to be an abundance of agents who will be answerable for the outcome in cases like Massacre.⁵⁸¹ The account of answerability which I have defended contains, therefore, no gaps.

7.3. Someone to blame

Recall the claim that I was not responsible for Julie's failure to get to school on time. We have seen one sense of 'responsibility' in which that claim seems false: I am *answerable* for Julie's failure to get to school on time to the extent that my actions or omissions causally contributed to its occurrence, that I was in control of my actions or omissions and that at the time of my causal contribution it was reasonably foreseeable that an outcome of that type or kind might occur. But after you have called me to answer for my conduct and I have explained that I took all reasonable steps to ensure that Julie would arrive at school on time—in short, when I have established that my conduct was in no way *culpable*—then there is another sense of 'responsible' in which it seems true that I am not responsible for her failure to get to school on time. This sense of responsibility has to do with blameworthiness. What a number of authors seem to mean by the Responsibility Gaps Thesis, then, is that certain AI systems such as the LAW in Massacre are capable of causing harms such that the degree of blameworthiness which can be allocated to any (group of) person(s) on our best theories falls short of how much blameworthiness should be allocated to those (groups of) person(s) according to our intuitions.⁵⁸² In this section, I show that this interpretation of the Responsibility Gaps Thesis is also false: there are no 'gaps' in blameworthiness.

⁵⁸¹ Granted, Kiener's 'abundance' is likely to encompass more agents, since he rejects any threshold for the causal significance of a contribution. But the 'abundance' which I have defended is more consequential, reflecting the fact that lots of agents bear a robust kind of answerability in cases like Massacre.

⁵⁸² E.g., Eggert (2023: 5): 'certain autonomous systems [may] act with such independence from human agents that harms they inflict on human persons cannot be traced back to a human violation of a human duty'; Matthias (2004: 175); Tigard (2020).

To begin with, a finding of answerability may seem to warrant, in certain circumstances, blame. For example, after I find you answerable for Julie's failure to get to school on time, I may also blame you if it turns out that you failed to take reasonable steps to ensure that she would get to school on time. I suggest that where an agent is answerable for some bad outcome and the agent causally contributed to the production of that outcome in a way which fell below some standard of reasonable care then blame may be warranted.⁵⁸³ Contrast, for example, the following two cases.

Reckless Driver: Driver S speeds around a busy town centre at 60 mph because she loves the speed. A Deliveroo employee, V, darts across the road from the restaurant to his bike. S tries to brake but collides with V, causing life-changing injuries. The accident would not have happened if S had been doing the speed limit.

Conscientious Driver: Driver T drives carefully around a busy town centre at 20 mph, keeping a careful look-out for pedestrians and cyclists. A Deliveroo employee, V, darts across the road from the restaurant to his bike. T tries to brake but collides with V, causing life-changing injuries.

Both S and T are answerable for the injuries to V. Their actions causally contributed to V's injuries and the possibility of causing injuries of that type and kind is a reasonably foreseeable consequence of driving around a busy town centre—even for T, as the empirical literature on no-fault accidents will attest to.⁵⁸⁴ Both S and T were also sensitive to reasons for and against their driving. But only S seems to be blameworthy for the outcome given her failure to take reasonable care. This illustrates the idea that a necessary condition for an agent to be

⁵⁸³ C.f., Feinberg (1970: 222): 'in the standard case of responsibility for harm, there can be no liability without contributory fault'.

⁵⁸⁴ Similarly, Williams' lorry driver is answerable for the child's death on the robust account of answerability which I have defended. Kiener's assumption that only strict moral answerability solves the paradox depends upon his misunderstanding of the standard control and epistemic conditions for moral responsibility (Kiener 2024: 362).

blameworthy for some bad outcome is that the agent's causal contribution to the production of that outcome was in some sense culpable. This prompts the question: in Massacre, is anyone blameworthy for the outcome?

I submit that there is nothing on the facts of cases like Massacre to suggest that anyone should be blameworthy for the outcome. Moreover, this fits with our account of blameworthiness because *a fortiori* no one is at fault. Recall the idealising assumptions that the LAW in Massacre is tried-and-tested, more reliable than the average human pilot at distinguishing civilians from military targets and deployed in a war which satisfies the conditions for being a just war. In that case, there is nothing necessarily negligent about the deployment of a LAW. One helpful way to illustrate this idea is put forward by Simpson and Müller,⁵⁸⁵ who point out that if a LAW operates within a reasonable tolerance level (e.g., it is at least as good as a human alternative at abiding within the limits of *jus in bello* conditions) and the war otherwise satisfies the conditions for being a just war, then the deployment of a LAW does not appear to be intrinsically culpable. In addition, let us assume that the programmer, seller and commander of the LAW take all reasonable care to protect against such an outcome as transpires in Massacre. In other words, these agents are more like the Conscientious Driver than the Reckless Driver. It therefore seems to me that our theories of blameworthiness get intuitively the right (null) results in cases like Massacre.

In sum, I have argued that lots of people are answerable for the outcome in Massacre but no one seems to be blameworthy. I have argued that the Responsibility Gaps Thesis is false on either conception of moral responsibility: there is no discrepancy between our best theories and our intuitions when it comes to answerability; nor is there any discrepancy between our intuitions about blameworthiness and our theory. But this prompts the following objection.

⁵⁸⁵ Simpson and Müller (2016).

Those who causally contribute to the production, sale and use of the LAW may not be blameworthy and therefore not liable to moral and legal censure like the Reckless Driver. But nor are they quite like the Conscientious Driver. They are still, after all, contributing to the production, sale and use of war machines which carry a certain kind of characteristic risk, risk which materialises in cases like Massacre. Surely, it seems, we want more than answers from these people. For one thing, we want reparations and compensation for the victims and their families, financial assistance in the reconstruction of hospitals and schools and meaningful steps to ensure that this kind of event should never happen again. How might we, in other words, ensure legal accountability for the outcome in Massacre? It is to this question that I turn in the next chapter.

8. Criminal and civil liability gaps

*Right the wrongs you made*⁵⁸⁶

This chapter considers a third and final interpretation of the Responsibility Gaps Thesis which some authors espouse, namely that there is some discrepancy between the (degree of) legal responsibility and the (degree of) moral responsibility which is attributable in cases like Massacre. Most legal systems distinguish between criminal and private law: criminal laws (usually) are enforced by the state, apply irrespective of whether a given perpetrator has agreed to be bound by them and are punishable by imprisonment and/or fines; private law regulates legal relationships between individual/group agents which may be initially voluntary; ‘involves the creation, alteration, and destruction of rights’⁵⁸⁷ and its usual remedy is a payment of money. At the international level, criminal and private law remedies may also be available following a legally recognised wrong, a truth which this chapter’s two-part analysis of international legal remedies available in cases like Massacre reflects.

Firstly, in **8.1**, I examine whether there is any discrepancy between what is in fact criminalised and who morally deserves punishment or who it would otherwise be instrumentally good to punish in cases involving LAWs.⁵⁸⁸ I conclude that there is not, in principle,⁵⁸⁹ any ‘gap’ when it comes to attributions of criminal responsibility. Secondly, in **8.2**,

⁵⁸⁶ Chapman (1989).

⁵⁸⁷ Turner (2019: 82). I follow Turner (*ibid*, n.1) in preferring the designation ‘private law’ to ‘civil law’ since the latter is ambiguous between legal systems that are founded upon codification and where judicial precedent does not play the same role as in common law systems.

⁵⁸⁸ I follow Tadros (2011) in taking there to be two classes of justifying aims of punishment. Retributivists ‘believe that criminal wrongdoers deserve to suffer and that the suffering of wrongdoers is intrinsically valuable’ (*ibid*: 26). Instrumentalists believe that ‘[p]unishment might be justified in the light of its instrumental rather than its intrinsic benefits’ (*ibid*: 28). I remain neutral as to the merits of either class.

⁵⁸⁹ Since this is a legal philosophy thesis, any claims about the existence or not of ‘gaps’ in legal responsibility attributions will contain an implicit ‘in principle’ caveat. This is because it is beyond my scope to reflect on the well-documented limitations ‘in practice’ concerning the enforceability of legal responsibility attributions, particularly in international criminal law. For an insightful overview of this latter topic: Adamu (2020).

I examine whether there is any discrepancy between the reparatory moral duties which people have in cases involving LAWs and the kinds of remedies which are actually enforceable under international private law. To anticipate, I will articulate a novel argument that those who participate in enterprises which produce, sell and use LAWs bear significant reparatory moral duties to the victims in cases like Massacre where the LAW causes an unpredictable lethal accident. This is because the risk of such accidents is a characteristic risk of those enterprises' activities. These significant reparatory moral duties are, however, not reflected in existing private law mechanisms applicable in cases like Massacre. This, I submit, is unsustainable given the ongoing development of LAWs and the ever-growing risk of cases like Massacre actually materialising. It is to the remedy of this state of affairs—this 'gap' in legal responsibility—that I turn in the next chapter.

8.1. Criminal responsibility for Massacre

It is commonly assumed in the AI ethics literature that there is some gap between what is in fact criminalised under international humanitarian law and who should morally be punished in cases like Massacre. For instance, in his recent book *Individual Criminal Responsibility for Autonomous Weapons Systems in International Criminal Law*, de Vries argues that 'in those cases [like Massacre] where no-one gave direct instructions, where [a LAW] goes beyond its instructions'⁵⁹⁰ then 'there is no form of criminal responsibility in international criminal law that can truly be applied'⁵⁹¹ which 'brings into question the ethicality of developing systems whose potential illegal actions cannot be addressed through criminal responsibility'.⁵⁹² De Vries concludes that 'either international criminal law will need to be altered to allow for accountability' in cases like Massacre or 'otherwise there will be no possibility within

⁵⁹⁰ De Vries (2023: 252).

⁵⁹¹ *Ibid*: 254.

⁵⁹² *Ibid*: 255.

international criminal law to hold anyone accountable for the actions of [LAWs].’⁵⁹³ In essence, I argue that this conclusion is undermotivated: there is no need, in principle, to alter international criminal law to ensure appropriate attributions of legal responsibility in cases involving LAWs.

Firstly, in **8.1.1**, I show that no one bears criminal responsibility for the outcome in Massacre under international law as it actually stands. I suggest that de Vries would agree with me. Secondly, in **8.1.2**, I show that, *pace* de Vries, this state of affairs is entirely appropriate given that in cases like Massacre it is neither intrinsically nor instrumentally good to punish anyone.⁵⁹⁴ Thirdly, **8.1.3** highlights that there is, in principle, no reason why blameworthy contributions to harms involving LAWs would not be punishable under the existing mechanisms of international criminal law. In such a short space, I do not purport to offer any conclusive arguments against the existence of ‘gaps’ in international criminal law in cases involving LAWs. But my arguments will suffice to motivate a shift in focus in **8.2**.

8.1.1. Criminal responsibility

Under international humanitarian law, there are two central ways to attribute individual responsibility for war crimes: *directly* and *indirectly*.⁵⁹⁵ *Direct responsibility* requires ‘the material elements’ of an offence to be ‘committed with intent and knowledge’.⁵⁹⁶ In certain cases, recklessness may suffice, which the ICRC terms ‘the attitude of an agent who, without being certain of a particular result, accepts the possibility of it happening’.⁵⁹⁷ But *a fortiori* no agent in Massacre intentionally, knowingly or recklessly contributes to the outcome’s

⁵⁹³ *Ibid*: 257.

⁵⁹⁴ Although punishment is arguably not the sole justificatory aim of international criminal responsibility attributions, it seems to be the primary one (Heinze 2020).

⁵⁹⁵ Crootof (2016: §II); HRW (2015).

⁵⁹⁶ Rome Statute, Arts.30(1)-30(3).

⁵⁹⁷ ICRC Commentary (1987: 994); *Prosecutor v. Delalić* [1998] Case No. IT-96-21-T at [437]-[439].

production in Massacre. As such, it is generally accepted that no one satisfies this mental condition for the attribution of direct responsibility in cases like Massacre.⁵⁹⁸

Indirect (or *command*) responsibility standardly holds that a military commander or civilian superior may be liable for a subordinate's crime if she (i) exercises 'effective [...] control' over a subordinate;⁵⁹⁹ (ii) 'knew or had reason to know' of the subordinate's actual/intended criminal acts;⁶⁰⁰ and (iii) 'failed to take the necessary and reasonable measures' to prevent or punish them.⁶⁰¹ Since the doctrine is premised on an individual's omissions or failure to fulfill a duty,⁶⁰² when evaluating (ii) and (iii), tribunals tend to apply something akin to a gross negligence standard, which involves 'reckless disregard of a legal duty'.⁶⁰³ A number of authors have argued for the attribution of legal responsibility for 'crimes' committed by LAWs to their military or civilian superiors on the basis of this doctrine.⁶⁰⁴

There are three principal threads of objections in the literature to the application of the doctrine of indirect responsibility to cases like Massacre. The first is that indirect responsibility is parasitic upon a direct responsibility attribution: superiors are responsible for failures to prevent only *chargeable criminal offences* by their subordinates.⁶⁰⁵ In other words, the commander can only be indirectly responsible if there was, in the first instance, a crime committed. In Massacre, however, that condition is not satisfied: LAWs, lacking the mental

⁵⁹⁸ For an overview of alternative international legal mechanisms for attributing responsibility directly: de Vries (2023: ch.7). See also: HRW (2012, 2015, 2018); McFarland and McCormack (2014).

⁵⁹⁹ Rome Statute, Art.28(a).

⁶⁰⁰ ICTR Statute, Art.6(3).

⁶⁰¹ ICTY Statute, Art.7(3). C.f.: Sierra Leone Statute 2002, Art.6(3); Additional Protocol I, Arts.86-87; *Prosecutor v. Delalić* at [346]. On the development of this customary rule: Cassese (2013: 182-187).

⁶⁰² Additional Protocol I, Art.86.1.

⁶⁰³ Bryan (1999: 1057); Cassese (2013: 53). For criticism of this approach: Bonafé (2009: 271).

⁶⁰⁴ See, e.g., Buchan and Tsagourias (2020); Margulies (2017); Jain (2016); Kraska (2021). Corn (2014) has articulated and defended a theory of 'procurement responsibility', according to which the doctrine of indirect responsibility should be applied to the military and civilian *procurers* of dangerous military technology since it is they who knowingly procured a dangerous weapon. I take no stand on the independent merits of this proposal here: since Massacre involves a tried-and-tested weapon, it is not clear that anyone knowingly procured a dangerous weapon in the requisite sense.

⁶⁰⁵ Mettraux (2009: 131-134).

element, cannot commit crimes.⁶⁰⁶ Even if that first problem can be surmounted,⁶⁰⁷ the second kind of objection is that it is difficult to see how the ‘effective control’ requirement could be met in the case of LAWs. The ICC Trial Chamber in *Bemba* confirmed that ‘effective control’ refers to ‘the material ability to prevent or repress the commission of the crimes or submit the matter to the competent authorities’.⁶⁰⁸ But, says Spadaro, there is no such ability when it comes to LAWs since their ‘superior’

‘would not hold any disciplinary power over them, would be unable to punish them and would also be unable to remit the case to the judicial authorities, given that machines cannot be prosecuted.’⁶⁰⁹

This does not seem to me a decisive objection given that there may be analogues for the exercise of disciplinary power, punishment or remission of the matter to the competent authorities in respect of LAWs. Burri, for example, has suggested that although LAWs, lacking sentience, may not be made to suffer, some of the instrumental goals of punishment may be achieved in cases like Massacre through destroying or reprogramming the LAW or even ‘[d]emanding that the machine apologize or provide some form of compensation to the families and dependents of its victims’.⁶¹⁰ To the extent that the justifying aims of ‘effective control’ are not limited to the intrinsic value of creating suffering, there seems to be no reason in principle why some of those aims should not be achievable in the context of LAWs.

But there is a third, more fundamental kind of objection to the application of indirect responsibility in cases like Massacre. This centres on the epistemic element for indirect

⁶⁰⁶ See, e.g., HRW (2015); McFarland and McCormack (2014); Spadaro (2023).

⁶⁰⁷ Some authors think that it can be (Buchan and Tsagourias 2020). For a counterargument: Spadaro (2023). My arguments in the next chapter for their legal personification would be a corollary for the attribution of legal responsibility to LAWs, whether criminal or civil.

⁶⁰⁸ *Bemba Gombo No.2* [2016] (ICC-01/05-01/08) at [183]; *Bemba Gombo No.1* [2009] (ICC-01/05-01/08) at [414]–[415].

⁶⁰⁹ Spadaro (2023: 12).

⁶¹⁰ Burri (2017: 180).

responsibility, namely that the superior knew or had reason to know of the subordinate's actual or intended criminal acts. But the LAW in Massacre is tried-and-tested. Overall, it operates more reliably in distinguishing legitimate from illegitimate targets than a human soldier. Of course, LAWs are to some extent intrinsically unpredictable. Moreover, LAWs may be 'not transparent in the way they function and provide no explanation for why they produce a given output'.⁶¹¹ But Corn points out that this is no different from the standard kind of autonomous agent which we habitually deploy on the battlefield, namely the human soldier.⁶¹² Commanders are not held criminally responsible for the unpredictable wrongful acts of human soldiers where these appear without warning because the commander's failure to foresee such acts is not blameworthy. For this reason, commentators object that the issue in cases like Massacre remains fundamentally unresolved: holding the commander responsible 'stretches individual blameworthiness to the outer limits of no-fault criminal responsibility'⁶¹³ and even threatens to undermine the moral legitimacy of international criminal law, which is premised on the principle of *nulla poene sine culpa*.⁶¹⁴ For the foregoing reasons, it is generally agreed that cases like Massacre do not fall within the ambit of the principal mechanisms under international law for allocating criminal responsibility. So far, so harmonious with de Vries. Is this state of affairs *right*?

8.1.2. Punishment is no good (in Massacre)

In this subsection, I argue that it is entirely appropriate that, in cases like Massacre, no one bears criminal responsibility under international criminal law as it currently stands. In essence,

⁶¹¹ ICRC (2019: 18).

⁶¹² Corn (2016).

⁶¹³ Amoroso (2020: 143).

⁶¹⁴ Crootof (2016: 1384); Seixas-Nuñez (2022b: 206). The principle that criminal prosecution requires culpable conduct is guaranteed under a number of human rights instruments (Universal Declaration of Human Rights, Art.11; Charter of Fundamental Rights of the European Union, Art.48[1]).

this is because it is neither intrinsically (8.1.2.1) nor instrumentally (8.1.2.2) good to punish anyone in such cases.

8.1.2.1. Retributivists

Retributivists believe that ‘[i]t is intrinsically valuable that offenders suffer in proportion to the gravity of the offence they have committed because that is what they deserve’.⁶¹⁵ There appear to be two intertwined reasons why an offender might deserve punishment for some conduct—in brief, punishment might be deserved in virtue of the nature of the conduct itself or in virtue of their character as it is expressed through that conduct.⁶¹⁶ I will suggest that neither of these reasons is applicable in cases like Massacre, whilst remaining neutral as to the truth or otherwise of any variant of retributivism.

Firstly, an offender might deserve punishment in virtue of certain kinds of conduct. There is some dispute as to which kinds of conduct might merit punishment. Gardner, for example, believes that ‘the only actions that deserved to be punished are both *wrongful* and *blameworthy*’,⁶¹⁷ where wrongfulness denotes that there has been a breach of a duty and blameworthiness in this context indicates that the breach of duty was neither justified nor excused. Dempsey, by contrast, has argued that breach of duty is not necessary for deserving punishment but makes a difference when it comes to standing to punish.⁶¹⁸ I take no stand on this debate here because it is generally agreed that punishment-deserving conduct should be blameworthy and, as we saw in the last chapter, given the stipulated idealising assumptions, no one’s conduct appears to be blameworthy in Massacre.

⁶¹⁵ Tadros (2011: 26); c.f. Berman (2024: 44). Retributivists also believe that ‘it is permissible for the state to ensure that offenders get what they deserve by punishing them’ (Tadros 2011: 26; Berman 2024: 44).

⁶¹⁶ A consequentialist might find that punishment is deserved in virtue of the good consequences that flow from its imposition. I will consider these kinds of arguments when I consider instrumentalist theories of punishment in 8.1.2.2.

⁶¹⁷ Gardner (2005, 2021).

⁶¹⁸ Dempsey (2023).

Secondly, an offender might deserve punishment in virtue of the kind of character which their conduct expresses. One way to think about this is as follows. According to Feinberg's expressive theory of punishment:

‘Punishment is a conventional device for the expression of attitudes of resentment and indignation, and of judgments of disapproval and reprobation, either on the part of the punishing authority himself or of those ‘in whose name’ the punishment is inflicted.’⁶¹⁹

Punishment, then, is justified only if those attitudes and judgments are justified. One account of the justification of such attitudes and judgments is Strawson's view that it is ‘the quality of others’ wills towards us, as manifested in their behaviour’⁶²⁰ which may or may not render them liable to justified attitudes of resentment and indignation and judgments of disapproval and reprobation.⁶²¹ For example, Strawson contrasts cases where I injure you accidentally or deliberately. The appropriate attitudes and judgments in each case depend, argues Strawson, on the quality of will which is expressed in my conduct. What kind of quality of will is expressed in Massacre? Recall the idealising assumptions that the LAW is tried-and-tested, more reliable than the average human pilot at distinguishing between civilian objects and military targets and deployed in a war which satisfies the conditions for being a just war. Indeed, the programmer, seller and commander of the LAW seem to be conscientious characters with no intention—quite the opposite—of causally contributing to such an outcome. Their behaviour, in other words, reflects a quality of will involving appropriate respect towards non-combatants on the opposing side, to the extent that anyone working in the production, sale and use of weapons like LAWs can exhibit such respect. Intuitively, I suggest, those who contribute to the

⁶¹⁹ Feinberg (1965: 400).

⁶²⁰ Strawson (1962[2008]: 15).

⁶²¹ *Ibid*: 5.

production, sale and use of the LAW in Massacre do not deserve punishment either for their conduct or for the character which their conduct expresses.⁶²²

8.1.2.2. Instrumentalists

Punishment might also be justified in virtue of its instrumental benefits.⁶²³ There are at least four kinds of instrumental benefits which might matter, individually or in combination, for the purposes of justifying punishment, which I call ‘paternalist’,⁶²⁴ reparation, incapacitation and deterrence views. According to paternalist views,⁶²⁵ part or all of the justification for punishment lies in the rehabilitative benefits which are received by the offender herself. Reparation views, by contrast, take the aim of punishment to be at least in part some form of valuable reconciliation between a wrongdoer and the victim and/or wider political community through the recognition by the offender that what she did was wrong by undergoing harsh treatment.⁶²⁶ Incapacitation views focus on the good of preventing offenders from committing further offences although, as Tadros points out, ‘[i]ncapacitation alone [...] is not punishment’⁶²⁷ since we can incapacitate a person without intending that they should suffer at all. Deterrence views come in two kinds. Special deterrence views take part of the justification for punishment to be that an offender may be discouraged or deterred from committing further crimes. General deterrence views concern the good of deterring other people from committing crimes by the punishment of offenders.⁶²⁸

⁶²² Otsuka (1998) defends the alternative ‘principle of avoidable blame’ as a diagnosis of why we respond to careful and careless accidents differently. But the point stands: cases like Massacre involve careful accidents. There is, as such, no blame to be attributed.

⁶²³ It is usually assumed that something over and above mere instrumental benefit, such as a duty to punish, is necessary for the justification of punishment to avoid using wrongdoers as means to an end or, indeed, to explain what is wrong with punishing the innocent. See Tomlin (2017).

⁶²⁴ Tadros (2011: 28).

⁶²⁵ Tadros (*ibid*) distinguishes between ‘pure paternalists’ who believe that ‘the only goods that are relevant to the justification of punishment are goods that are received by the offender herself’ (e.g., Hampton 1984) and ‘less pure’ paternalists who are open to other goods mattering too (e.g., Morris 1981).

⁶²⁶ This is one way of reading Duff’s communicative theory of punishment (Duff 2001; c.f. Tadros 2011: 29, n.13).

⁶²⁷ Tadros (2011: 29).

⁶²⁸ For criticism of general deterrence views: Quinn (1993).

I take no stand on the merits of any of these instrumentalist views as none of them apply in Massacre. It is clear that in cases like Massacre those who have contributed to the production, sale and use of the LAW have no need of rehabilitation. Moreover, lethal accidents of this nature do not seem to call for the recognition of wrongdoing through undergoing hard treatment; rather, as I argue in 8.2, reparations should ensure that those who participate in enterprises which produce characteristic risks for others pay the costs when those characteristic risks materialise. Finally, neither incapacitation nor deterrence is warranted as there is no need to encourage or otherwise engender a change in the behaviour of the contributors to the production, sale and use of the LAW or in the behaviour of people more generally. In sum, whilst the considerations advanced in this section do not purport to be decisive, they should make us sceptical of claims that international criminal law as it currently stands contains ‘gaps’. On the contrary, I have suggested that *pace* de Vries, in cases like Massacre no one deserves punishment, nor would punishment serve any instrumental good.

8.1.3. Criminal responsibility revisited

One might interject at this point: do other kinds of cases involving harms caused through the use of LAWs generate any morally problematic ‘gaps’ in international criminal responsibility? Consider, by way of illustration, the following case:

Malicious Massacre: Everything is the same as in Massacre except that this time the LAW was programmed by a malicious programmer to target civilians.

As we saw in the last chapter, the programmer would be not just answerable but blameworthy for the outcome in Malicious Massacre. Moreover, on each of the theories of the moral justification of punishment glossed *supra*, punishment (in some form) of the programmer would seem to be justified. Would she incur liability under international criminal law?

In brief, yes. Consider the mechanism of direct responsibility attribution. Clearly, the mental element of intentionally, knowingly or recklessly contributing to the perpetration of a crime is present in Malicious Massacre. Is there anything about the nature of the programmer's causal connection to the outcome which might jeopardise the attribution of responsibility directly? Per the Appeals Chamber of the ICTY in *Tadić*, we must understand direct criminal responsibility as involving the 'physical perpetration of a crime by the accused himself, or a culpable omission of an act mandated by a rule of criminal law'.⁶²⁹ This definition has been confirmed in subsequent decisions of the ICTY,⁶³⁰ ICTR⁶³¹ and other tribunals.⁶³² De Vries has highlighted how programming (or commanding) a LAW to commit an atrocity such as Malicious Massacre would straightforwardly satisfy this definition because:

'[The LAW] in such [a] case does not actually operate in an autonomous manner and should be considered [...] as any tool used in the commission of a crime. Programming or commanding [a LAW] would be no different than firing a gun or placing a landmine [...] Therefore, although LAWs function in a more sophisticated manner, [...] the regular concept of commission would be applicable. It would be the programming in of the committing of the crime that would constitute the *actus reus* and should therefore be seen as any other form of direct perpetration.'⁶³³

I submit that, although there is no precedent on this particular issue, cases like Malicious Massacre would fall within the ambit of existing mechanisms for the attribution of criminal responsibility under international law. So, under international law, there seems to be no

⁶²⁹ *Prosecutor v. Tadić* [1999] Case No.it-94-1-a at [188].

⁶³⁰ *Prosecutor v. Stanislav Galić* [2003] Case No.it-98-29-t at [168]; *Prosecutor v. Limaj et al.* [2005] Case No.it-03-66-t at [509].

⁶³¹ *Prosecutor v. Nahimana* [2007] Case No.ictr 99-52-a at [478].

⁶³² C.f. *Prosecutor v. Brima* [2006] (scsl-04-16-t) 35: 'An individual can be said to have 'committed' a crime when he or she physically perpetrates the relevant criminal act or engenders a culpable omission in violation of a rule of criminal law'.

⁶³³ De Vries (2023: 187).

discrepancy between what is in fact criminalised and the intrinsic or instrumental goods of punishment in cases involving LAWs. When it comes to criminal responsibility at least, the Responsibility Gaps Thesis is therefore false. In the next section, I consider whether there might nonetheless be some kind of ‘gap’ when it comes to the kinds of remedies which are available to victims under international private law as it currently stands. I show that there is a ‘gap’. In the next chapter, I propose a novel and timely solution to the problem which I have revealed.

8.2. Keeping things civil

This section articulates and defends a novel argument to the effect that all those who participate in enterprises which produce, sell and deploy LAWs bear reparatory moral duties in cases like Massacre (8.2.1). These duties are, however, not reflected in international private law (8.2.2). It is to the development of a novel solution to this ‘gap’ in private law remedies in cases like Massacre that I turn in the next chapter. Although the term suffers from vagueness, I will understand an ‘enterprise’ to have two core features: it is an *institution*—broadly, an organised collective, structured by rules—for the *exchange of goods and services*. An enterprise need not make a profit or even aspire to; it can be a public or a private body.⁶³⁴

8.2.1. Reparatory moral duties for Massacre

My novel contribution to the literature at this stage in the dissertation is to offer an alternative way to think about the issue of responsibility in cases like Massacre. As we have seen, the AI ethics literature—and the LAWs literature in particular—has long been preoccupied with the Responsibility Gaps Thesis. But I have shown that three popular interpretations of that thesis are false. This prompts the question: what might be motivating this enduring, inchoate concern about responsibility for LAWs in cases like Massacre? In this section, I want to suggest that one thing which might be going on is as follows. The production, sale and use of LAWs involve

⁶³⁴ I adopt this conception of an ‘enterprise’ since it aligns with jurisprudence across the common law world, which I discuss in the next chapter. See also Deakin (2003); Tan (2022). For criticism: Morgan (2012).

enterprises—technology companies, military suppliers and militaries themselves—which create a characteristic risk as part of their legitimate operation.⁶³⁵ When that characteristic risk materialises, it can lie where it falls on innocent civilians and their communities or it can be absorbed by the enterprises which created it. In this section, I argue for the latter approach.

No one so far has defended this kind of view in the AI ethics literature. It is also an original contribution to the political philosophy literature on reparatory duties. Firstly, I show that Miller’s ‘connection theory’⁶³⁶ of reparatory duties does not seem to get the right answers in cases like Massacre (8.2.1.1). Next, I show that Butt’s amendment to that theory, namely the ‘beneficiary principle’,⁶³⁷ will also not suffice (8.2.1.2). A tempting alternative approach is to view the problem through the lens of structural injustice but this turns out to be inadequate too (8.2.1.3). Finally, I draw on Keating to outline a novel ‘enterprise liability’⁶³⁸ principle for the attribution of significant reparatory moral duties in cases like Massacre: those who participate in enterprises which produce, sell and use LAWS bear reparatory moral duties when the characteristic risks of those activities materialise in cases like Massacre (8.2.1.4).

8.2.1.1. The connection theory

Miller sets out to address what he calls ‘the problem of remedial responsibility’,⁶³⁹ that is, given some ‘bad situation’⁶⁴⁰ involving ‘suffering and deprivation’,⁶⁴¹ the problem of identifying who has ‘a special obligation to put the bad situation right, in other words to be picked out, either individually or along with others, as having a responsibility towards the deprived or suffering

⁶³⁵ I do not endorse any particular concept of group agency in this dissertation. When I refer to an ‘enterprise’, that can be interpreted (non-exhaustively) as shorthand for a collection of individuals or as an entity more than the sum of its parts.

⁶³⁶ Miller (2002).

⁶³⁷ Butt (2007, 2014).

⁶³⁸ Keating (2023).

⁶³⁹ Miller (2002: 453).

⁶⁴⁰ *Ibid.*

⁶⁴¹ *Ibid.*

party that is not shared equally among all agents.’⁶⁴² Miller identifies four relations which may connect an agent to a bad situation. The first two are ‘backwards-looking’ principles—that is, focused on the question, ‘Who is responsible for bringing this bad situation about?’⁶⁴³—which appeal to ‘causal’ or ‘moral responsibility’ for the occurrence of the condition. The second two are ‘forwards-looking’ principles—that is, focused on the question, ‘Who is best placed to put it right?’⁶⁴⁴—which appeal to capacity for remedying the condition and communal obligations to the affected parties.⁶⁴⁵ Concluding that no single relation encompasses all those who should remedy any given situation, Miller expounds what he calls ‘the connection theory’ of responsibility according to which any of the principles which he considers may establish a sufficiently strong link between an agent and a bad situation to generate a special obligation to put the bad situation right. I briefly gloss these alternatives to show that this standard account of reparatory moral duties does not get the right answers in Massacre.

The first principle which Miller considers is the causal principle according to which responsibility for remedying a bad situation should be distributed among agents in proportion to their causal contribution to that situation’s occurrence. However, in cases of non-culpable causation of injury like Massacre—or, in Miller’s words, ‘innocent causation’⁶⁴⁶—Miller allows that ‘bare causation’ may play only a ‘minor part in distributing responsibilities’.⁶⁴⁷ For example, as we saw in the case of Williams’ lorry driver, if I cause an accident Miller considers it ‘appropriate that I should apologise, or at least express regret for what has happened’,⁶⁴⁸ but not that I should bear substantial reparatory moral duties. In Massacre, then, those who causally contribute to the outcome’s production should apologise or express regret on Miller’s view. But

⁶⁴² *Ibid.*: 454.

⁶⁴³ *Ibid.*: 460.

⁶⁴⁴ *Ibid.*

⁶⁴⁵ On the distinction between backwards- and forwards-looking principles: van de Poel (2011). The distinction is somewhat artificial but I take no stand as to its ultimate merits here.

⁶⁴⁶ Miller (2002: 458).

⁶⁴⁷ *Ibid.*

⁶⁴⁸ *Ibid.*

this might seem a rather tepid response if we reflect, as I propose to in this section, on the fact that those contributors are part of an enterprise which foreseeably creates risks to others, albeit not in a blameworthy way.

A similar problem attends the second principle which Miller considers, namely the moral responsibility principle. Miller distinguishes between two concepts of moral responsibility. The ‘narrow’ concept, which corresponds to what I have called blameworthiness, generates substantial reparatory moral duties: ‘A should be held remedially responsible for P’s condition insofar as he is *morally responsible* for its occurrence, in the sense that carries with it ascriptions of fault and blame.’⁶⁴⁹ That concept, as we have seen, carries little moment in the ascription of reparatory moral duties in cases like Massacre. The ‘broader’ concept corresponds roughly to what I have called answerability⁶⁵⁰ but Miller acknowledges that it is not standardly taken to generate any significant reparatory moral duties by itself.⁶⁵¹

The third and fourth principles emerge at a point in the argument when Miller decides to turn from backwards-looking modes of identifying remedial duty-bearers to forwards-looking ones. The third principle is the principle of capacity according to which special obligations to remedy a bad situation ‘ought to be assigned according to the capacity of each agent to discharge them’,⁶⁵² a principle which incorporates ‘the *effectiveness* of different agents in remedying the situation’ and ‘the *costs* they must bear’ in so doing.⁶⁵³ The fourth is the principle of ‘community’, according to which members of a shared community may owe special obligations to aid any member of the relevant community who is in need.⁶⁵⁴ Neither of these principles seems to appropriately identify those who participate in enterprises which

⁶⁴⁹ *Ibid.*

⁶⁵⁰ *Ibid.*: 459.

⁶⁵¹ *Ibid.*: 460.

⁶⁵² *Ibid.*: 461.

⁶⁵³ *Ibid.*

⁶⁵⁴ *Ibid.*: 462.

create characteristic risks of events like Massacre as bearers of any kind of special responsibility for the outcome when those risks materialise. Moreover, the capacity and community principles might not seem to capture the distinctive moral urgency of Massacre. Indeed, corrective or backwards-looking principles arguably call more pressingly on our consciences.⁶⁵⁵ Perhaps an alternative backwards-looking principle may assist in the ascription of reparatory moral duties to those who participate in enterprises for the production, sale and use of LAWs.

8.2.1.2. The beneficiary principle

One such principle has attracted widespread support in the climate change and historical injustice literatures, asserting that ‘agents can acquire rectificatory obligations through involuntarily benefiting from acts of injustice’.⁶⁵⁶ Call this the ‘beneficiary principle’. A core idea is that the receipt and retention of benefits of injustice would itself be blameworthy such that those who benefit from injustice may be required to disgorge those benefits in compensation.⁶⁵⁷ There are different ways to think about this. Gosseries describes the idea of retaining the benefits of injustice as that of being a ‘morally objectionable free rider’ on injustice.⁶⁵⁸ Barry and Wiens examine how receiving and retaining benefits can sustain wrongful harm, arguing that duties are acquired when ‘receipt and retention of the benefits of wrongdoing contribute to the persistence of wrongful harm suffered by the victim’.⁶⁵⁹ Butt considers that ‘we would not be in good standing as moral agents if we condemned the wrongdoing but refused to do anything in relation to its persisting effects’.⁶⁶⁰ I do not come

⁶⁵⁵ Mills (2015); Butt (2020, 2021); Tan (2007).

⁶⁵⁶ Butt (2014: 446); Caney (2006: 471); Meyer and Roser (2010: 252–53); Page (2011: 420–21, 2012).

⁶⁵⁷ Butt (2007, 2014); Page (2012).

⁶⁵⁸ Gosseries (2004: 47).

⁶⁵⁹ Barry and Wiens (2016: 2).

⁶⁶⁰ Butt (2020: 592; 2007).

down in favour of any particular underlying rationale for the principle but rather ask: what, if anything, might the beneficiary principle imply in cases like Massacre?

The beneficiary principle suffers from three weaknesses in its application to cases like Massacre, which I call the problems of identification, specification and enforceability.⁶⁶¹ Firstly, identification: it is not clear in principle how to identify the beneficiaries of injustice in such cases. Given the idealising assumptions, the production, sale and use of LAWs do not involve unjustified risk to, or infringement of, the rights of others. So, we cannot simply say that those who earn their salaries in those processes have benefited from injustice and should therefore disgorge the fruits.⁶⁶² Further, no one appears to benefit from the particular outcome in Massacre. No one is better off, for instance, if Massacre occurs or not.

Secondly, the problem of specification. Even if we could identify those who benefit from the injustice, the beneficiary principle provides for potentially very open-ended compensatory duties.⁶⁶³ Indeed, some authors suggest that the “‘contagion of injustice’ actually calls into question nearly all existing titles to property, such that moral agents should vacate their titles and permit a wide scale fresh distribution of property’”.⁶⁶⁴ Substantial reparatory moral duties are a good thing; under-specification of the extent of reparatory moral duties is not.⁶⁶⁵

Thirdly and relatedly, there are deep questions about the enforceability of the beneficiary principle. When he considers this question, Butt seems content to grant that the beneficiary principle may in fact generate reparatory moral duties which are not apt for legal

⁶⁶¹ Butt (2020) expands on three analogous difficulties for his theory generally.

⁶⁶² Although it may be a kind of *structural* injustice: see 8.2.1.3.

⁶⁶³ The problem becomes particularly acute if we expand the scope of the beneficiary principle to accommodate the way in which injustice may be (re)produced through social structures. See 8.2.1.3.

⁶⁶⁴ Butt (2020), citing Goodin (2013).

⁶⁶⁵ For alternative responses to this problem of specification: Haydar and Øverland (2014); Barry and Wiens (2016).

enforcement.⁶⁶⁶ Whilst a non-enforceable principle might still be of political significance, I submit that in cases like Massacre the scale of civilian harm calls for enforceable, specific reparatory moral duties with identifiable duty-bearers. I do not purport to have offered any decisive considerations against the applicability of the beneficiary principle in cases like Massacre. But the considerations advanced so far might prompt us to consider whether there is an alternative way to conceptualise injustice in such cases, an alternative lens which I will characterise as ‘structural’.

8.2.1.3. Responsibility for structural injustice

In her posthumously published book *Responsibility for Justice*, Young articulated and defended a theory of structural injustice according to which those who participate by their actions or omissions in the production and reproduction of structural injustice bear special responsibilities to work for its remedy, even if their participation is non-blameworthy.⁶⁶⁷ This might at first blush seem to offer a way to allocate reparatory moral duties in Massacre, one which captures the collective, non-blameworthy causation of harm and the way in which the benefits of a particular kind of social structure—one which legitimises a certain level of civilian harm in war—are distributed through institutions which produce, sell and use LAWs. I will show that the structural injustice framework suffers, however, from several weaknesses which motivate an alternative approach.

Structural injustice has three core features. Firstly, structural injustice involves systematic oppression and systematic privilege.⁶⁶⁸ Indeed, Young (2011) defines structural injustice as that phenomenon which

⁶⁶⁶ Butt (2020); c.f. Dunham and Lawford Smith (2017).

⁶⁶⁷ Modern theorising about structural injustice is usually traced to Young’s work on the subject in a series of articles in the 2000s and her posthumously published book (Young 2004, 2006, 2011).

⁶⁶⁸ I gloss over a debate in the literature concerning the distinction between structural *harm*—where social processes put people under systematic threat of domination or deprivation of the means to develop and exercise their capacities—and structural *injustice*—where those social processes at the same time enable others to dominate

‘exists when social processes put large groups of persons under systematic threat of domination or deprivation of the means to develop and exercise their capacities, at the same time that these processes enable others to dominate or to have a wide range of opportunities for developing and exercising capacities available to them.’⁶⁶⁹

The production, sale and use of LAWs seem to satisfy this definition. As we have seen, these social processes create a characteristic risk of cases like Massacre, placing large groups of persons under a systematic threat of deprivation of life and security of person. At the same time:

‘The global military AI market is growing quickly – it was estimated to be worth just under US\$9.2 billion in 2023 and is projected to at least double in the next five years [...] This, in turn, drives growth in the autonomous weapons industry, a market that is forecast to grow from an estimated US\$13.3 billion in 2022 to US\$21.8 billion by 2027 [...] The US and China are currently the dominant drivers of this market, ahead of other states like the UK, Israel, France and Russia in accelerating military AI rollout.’⁶⁷⁰

So, those social processes which place some groups of persons under systematic threat make others rich and powerful.⁶⁷¹

Secondly, structural injustice involves the ‘accumulated outcomes of the actions of [...] masses of individuals, enacting their own projects, often uncoordinated with many others’.⁶⁷² As a result, structural injustice is epistemically difficult to ‘trace’ to the causal contributions of individual actors. Similarly, the production, sale and use of LAWs generate injustices like

or enjoy abundant opportunities to develop and exercise their capacities (Wolff and Mantouvalou 2024: 14-19). Browne (2025: 40-41) focuses on structural harm; I agree with Wolff and Mantouvalou (2024: 16) that ‘structural injustice has an urgency and potency that structural harm lacks’.

⁶⁶⁹ Young (2011: 52).

⁶⁷⁰ Schwarz (2025: 238).

⁶⁷¹ Structural injustice is to that extent a feature of the wider military-industrial complex.

⁶⁷² Young (2011: 96).

Massacre which are epistemically difficult to trace to individual causal contributions. In part, this is due to the use of advanced machine learning technology, which is notoriously opaque.⁶⁷³ But ‘untraceability’ is also a perennial feature of international commercial relationships, as Young explored in her seminal analysis of ‘sweatshop labour’ as structural injustice.⁶⁷⁴ In the structural injustice literature, this difficulty of tracing individual causal contributions is one (pragmatic) motivation for a shift in focus from backwards-looking responsibility attributions to forwards-looking responsibility for remedying the structural conditions which enable the (re)production of structural injustice.⁶⁷⁵ But, as I develop below, cases like Massacre seem to call for more targeted, backwards-looking interventions than this structural lens will allow.

Thirdly, structural injustice ‘occurs as a consequence of many individuals and institutions acting to pursue their particular goals and interests, for the most part within the limits of accepted rules and norms.’⁶⁷⁶ Young illustrates this concept using two central examples. The first is the fictional story of Sandy, a low-income single-mother with two children who makes normal, proper and expected decisions within the housing market—as do the other characters in the story—but winds up on the brink of homelessness. The second is about sweatshop labour and how ordinary consumer decisions collectively result in alarming patterns of domination and deprivation.⁶⁷⁷ Similarly, the production, sale and use of LAWs, provided they meet the requirements of international humanitarian law, result from individuals and institutions pursuing goals within the limits of accepted rules and norms.⁶⁷⁸

I have suggested that the social process involving the production, sale and use of LAWs—and, arguably, the wider military industry—is a context of structural injustice. For

⁶⁷³ Browne (2023, 2025); Knight (2017); Nielly (2020); von Eschenbach (2021).

⁶⁷⁴ Young (2004).

⁶⁷⁵ Young (2011: ch.4).

⁶⁷⁶ *Ibid*: 96.

⁶⁷⁷ Young (2011: 125-30).

⁶⁷⁸ For an excellent critique of the adequacy of international humanitarian law for ensuring the moral permissibility of LAWs: Eggert (2023: 3).

Young, contexts of structural injustice engage the ‘social connection model’ of responsibility according to which all agents who are ‘connected’ to structures which produce injustice share ‘political’ responsibility (as opposed to moral or legal responsibility)—a nonblameworthy, forwards-looking and shared responsibility—to work collectively to challenge that structural injustice.⁶⁷⁹ But there are three reasons why a structural injustice analysis of Massacre is not enough. First and foremost, we need a response to *this harm*. Massacre involves a significant harm which materialises as a characteristic risk of the enterprises which produce, sell and use LAWs. For the people who participate in those enterprises to let the loss lie where it falls upon the immediate victims and their communities and work for structural change would seem to leave out something important—namely, as I contend in the next subsection, significant reparatory duties owed to the victims and their families. Secondly, the social connection model, because it seeks to move beyond a simplistic ‘victims’ and ‘perpetrators’ model of injustice to a model of responsibility as non-blameworthy, shared and forwards-looking, neglects some important features of the situation in Massacre. In particular, some enterprises have created the risk of Massacre as a characteristic risk of their activities; intuitively, those enterprises should absorb the cost of their activities rather than let it lie where it falls on the victims in Massacre. There is space in our theories of justice to move beyond a ‘victims’ and ‘perpetrators’ dichotomy; but that does not render otiose the identification of those who bear special reparatory moral duties for a given harm. Indeed, backwards-looking considerations arguably carry a distinctive moral urgency⁶⁸⁰ and motivate wider compliance.⁶⁸¹ Thirdly, the structural injustice lens leaves out an important feature of Massacre, namely unfairness. Those enterprises for the production, sale and use of the LAW have foisted the characteristic costs of their activities on those whose bad luck it is to get in the way. There is something intuitively *unfair*

⁶⁷⁹ Young (2011: ch.4). For an alternative account of the appropriate response to structural injustice: Zheng (2019).

⁶⁸⁰ Mills (2015); Butt (2020: 595).

⁶⁸¹ See Tan (2007: 287): ‘most ordinary and well-meaning people are more likely to take action to correct their past transgressions than to act as required by [forwards-looking] justice’.

and even *exploitative* about enterprises which behave in this way. This prompts me to consider the following approach to ascribing reparatory moral duties in cases like Massacre.

8.2.1.4. Enterprise liability

In July 2025, Sir Mark Rowley, head of the Metropolitan Police, argued that football clubs should pay towards the £70 million cost of policing their matches in the UK. At its heart, his case was that those (footballers in clubs) who participate in an enterprise (football matches) which poses a characteristic risk of some kind of harm (aggression at matches) should bear some or all of the cost of mitigating that risk, even if they would not be blameworthy for the harms should they occur.⁶⁸² I suggest that such a principle applies in cases like Massacre to those who participate in enterprises for the production, sale and use of the LAW: those agents should bear some or all of the cost of mitigating the risks which their enterprise characteristically generates, including through compensating the victims in cases like Massacre. Importantly, neither Sir Mark Rowley nor I are suggesting that the enterprise at issue is in any way creating unjustified risks. The characteristic negative side-effects of football matches and—given the stipulated idealising assumptions—the production, sale and use of LAWs may be entirely outweighed by the positive effects of the enterprise(s). Nonetheless, those who participate in that enterprise should pay for its negative side-effects. In cases like Massacre, I submit, significant reparatory moral duties may therefore attach to those who contributed to the production, sale and use of the LAW. In principle, these are legally enforceable duties but, as I discuss in the next subsection, the reparatory duties which apply under existing private law mechanisms are insufficient.

Before that, I wish to respond to three kinds of criticisms which one might make of my proposal that enterprise liability is engaged in cases like Massacre. Firstly, one might challenge

⁶⁸² Fatima (2025).

the persuasiveness of enterprise liability as a rationale for the ascription of reparatory moral duties at all. Indeed, whilst there is general agreement in the literature that various *policy* bases might warrant the ascription of reparatory duties for harms to those who participate in enterprises which create those harms as a characteristic risk of their operations,⁶⁸³ there is much less accord as to the existence of a *principled* basis for such ascriptions. One exception is Keating and it is upon his account of the fairness of enterprise liability that I wish to draw.⁶⁸⁴ Keating has argued that unfairness is the ‘moral nerve’⁶⁸⁵ of the wrong which enterprise liability engages with:

‘when an enterprise imposes the costs of unavoidable harms characteristic of its activities upon those that it happens to harm, the enterprise wrongly exploits those victims for its own gain. Repairing harm so inflicted erases that exploitation.’⁶⁸⁶

The idea is that nonblameworthy contribution to the causation of harm is not the wrong. Rather, the wrong lies in ‘foisting the cost of a harm that is an unavoidable byproduct of the activity onto the victim instead of taking responsibility for the harm and repairing the injury inflicted’.⁶⁸⁷ So, enterprise liability would seem an apt principle to respond to the collective, nonblameworthy but systematically unfair distribution of harm and benefit in cases like Massacre.

The second kind of concern which one might have regarding enterprise liability relates to the *kinds* of duties triggered. Granting that fairness is at the heart of enterprise liability, why should the duties triggered be considered backwards-looking, specific and potentially significant, as seemingly required in cases like Massacre? To answer this concern, we need to

⁶⁸³ For a helpful summary, see Keating’s (2023: 278-281) discussion of the ‘loss-spreading’ and ‘accident-prevention’ justifications for enterprise liability.

⁶⁸⁴ *Ibid.* ch.7.

⁶⁸⁵ *Ibid.* 281.

⁶⁸⁶ *Ibid.*

⁶⁸⁷ *Ibid.*

attend to the idea of exploitation in Keating's account of fairness. In general, '[t]o exploit someone is to take unfair advantage of them [that is,] to use another person's vulnerability for one's own benefit.'⁶⁸⁸ Fairness, on Keating's account, is therefore not a purely distributive notion. Rather, it involves righting a specific wrong which has occurred, namely the exploitation of 'harming-without-repairing' where the harm occurred as a result of characteristic risks of an enterprise.⁶⁸⁹ It is for this reason that Keating views enterprise liability as constituted by three commitments:

'(1) The costs of those accidents that are characteristic of an enterprise should be absorbed by the enterprise as an operating expense, not left on those whose bad luck it is to get in the enterprise's way;

(2) The costs of enterprise-related accidents should not be concentrated either on the victim who originally suffered the injury or on the particular agent who inflicted the injury; and

(3) The costs of such accidents should instead be distributed among those who benefit from the imposition of the enterprise's risks.'⁶⁹⁰

(1) arises from the backwards-looking idea of exploitation; (2) and (3) offer a framework for specifying potentially significant reparatory moral duties.

Thirdly, one might challenge the extent of those duties. Like the beneficiary principle, does enterprise liability risk going too far? Priest, for instance, has objected that enterprise liability fails as a rationale for the ascription of legally enforceable duties because its internal logic calls for unlimited liability and leads to unworkable forms of insurance.⁶⁹¹ But this seems

⁶⁸⁸ Zwolinski, Ferguson, and Wertheimer (2022).

⁶⁸⁹ Keating (2023: 281).

⁶⁹⁰ Keating (2023: 272).

⁶⁹¹ Priest (1985, 1987); Epstein (1985), discussed at Keating (2023: 275-6).

to be because Priest fails to appreciate the core idea of fairness in the enterprise liability logic. As Keating argues, the ‘responsibility ends when the risks of the enterprise cease to be distinctive [that is, characteristic] and merge into the general risks of social life.’⁶⁹² Whilst a full analysis of the notion of a characteristic risk would take us too far afield,⁶⁹³ it suffices for my purposes that those harms which are reasonably foreseeable consequences of the activities which individuals undertake as part of a given enterprise fall within the scope of enterprise liability. For example, in *Massacre*, the outcome is a characteristic risk of enterprises which produce, sell and deploy autonomous weapons of war. As such, under the enterprise liability rationale which I have defended, those who participate in such enterprises bear reparatory moral duties to compensate the victims. Let us now examine the extent to which these reparatory moral duties are legally enforceable under existing private law mechanisms.

8.2.2. Private law remedies

In his book *Robot Rules*, Turner surveys the two principal sources of private law obligations relating to AI:⁶⁹⁴ contract—which is based on agreement—and civil wrongs—viz., the infringement of the legal rights of one party. Within the law of contract, liability can arise through the formation of a contract or through the operation of insurance (8.2.2.1). Within civil wrongs, liability can arise in a number of different ways but the principal ones considered in the literature are negligence (8.2.2.2) and strict liability (8.2.2.3).⁶⁹⁵ I briefly survey the options and highlight a lacuna.

⁶⁹² Keating (2023: 276).

⁶⁹³ *Ibid*: ch.2.

⁶⁹⁴ Turner (2019: 83); c.f. Burrows (2017).

⁶⁹⁵ Turner (2019) also considers a no-fault compensation scheme for AI harms styled on New Zealand’s system for compensating road accidents. But he offers convincing reasons why this appears unworkable at a ‘scaled up’ level.

8.2.2.1. Contractual remedies

It has been suggested that those who produce and sell AI systems such as LAWs might voluntarily assume responsibility for harms caused by those systems.⁶⁹⁶ For example, in 2015, the CEO of Volvo announced that the company would accept liability for harm caused by its cars when they are operating autonomously.⁶⁹⁷ Turner suggests that Volvo could have thereby concluded a contract, drawing an analogy with the seminal English case *Carlill v Carbolic Smoke Ball Company*,⁶⁹⁸ where the company's boast that it would pay £100 to anyone who used their product and was not cured of 'flu was contractually binding.⁶⁹⁹ Alternatively, insurance represents a specific type of contract law in which the insurer agrees to pay money or undertake steps to compensate the insured if certain events occur. US Judge Karnow has suggested that an insurance scheme is the best way to deal with liability for no-fault accidents involving AI systems⁷⁰⁰ and in 2018 the UK Parliament enacted the Automated and Electric Vehicles Act which extends the compulsory vehicle insurance scheme to automated vehicles. One might worry about the first contractual option given its dependence on the voluntary agreement by companies to do something which they are independently morally required to do; one might worry about the role of insurance since it remains parasitic upon an underlying finding of liability for harm due to AI.⁷⁰¹ It is therefore to the investigation of whether some such underlying liability exists that I turn in the next subsection.

⁶⁹⁶ See, e.g., Kiener (2024, 2025).

⁶⁹⁷ Korosec (2015).

⁶⁹⁸ [1892] EWCA Civ 1.

⁶⁹⁹ There are important disanalogies: Carbolic used the claim as part of their marketing strategy and Carlill relied on it. Volvo used the claim to ask for regulatory clarity and there does not seem to have been reliance. Thanks to Grieman for helpful discussion.

⁷⁰⁰ Karnow (1996).

⁷⁰¹ Turner (2019: 116).

8.2.2.2. Liability for civil wrongs: negligence

Negligence is defined as conduct which fails to conform to a required standard.⁷⁰² That required standard is usually defined in terms of a duty of reasonable care, as we saw in the UK context in the seminal case of *Donoghue v Stevenson*:⁷⁰³ ‘you must take reasonable care to avoid acts or omissions which you can reasonably foresee would be likely to injure your neighbour’,⁷⁰⁴ where ‘neighbours’ were defined as ‘persons who are so closely and directly affected by my act that I ought reasonably to have them in contemplation’.⁷⁰⁵ Similar provisions apply under other legal systems.⁷⁰⁶ The obvious problem with enforcing reparatory moral duties in cases like Massacre via the mechanism of negligence is that *a fortiori* no agent’s conduct has fallen below a standard of reasonable care. In response to this problem, two approaches are notable: abandon the fault requirement, or adjust it. I do not purport to offer any decisive objections in so short a space to either kind of approach but I will suggest that the approach which I defend in the next chapter is to be preferred because it better satisfies the Dworkinian criteria of soundness and fit with existing law which Part One defended as criteria for assessing the legitimacy of new legal propositions.

One example of the former kind of approach is Crotoft’s proposal that we create a ‘war torts’ regime on which states bear ‘strict liability’ for LAWs-associated harms.⁷⁰⁷ Crotoft argues that such a regime would enable us to treat not only ‘morally blameworthy harms’—

⁷⁰² Turner (2019: 84); Nolan and Davies (2017: 934).

⁷⁰³ [1932] A.C. 562

⁷⁰⁴ *Donoghue v Stevenson* at 580-581.

⁷⁰⁵ *Ibid.*

⁷⁰⁶ Turner (2019: 85). Under the German Civil Code, S.823: ‘A person who, intentionally or negligently, unlawfully injures the life, body, health, freedom property or another right of another person is liable to make compensation to the other party for the damage arising from this’ (cited in Turner 2019: 85, n.19). Analogous provisions seem to apply in China too: Tort Law of the People’s Republic of China, 2009 (cited in Turner 2019: 85, n.20). France is somewhat different: it has been suggested that ‘in cases of personal injury and property loss, strict liability is the rule and fault liability the exception’ (Dam 2013: 51); the French Civil Code, Art.1382 provides that ‘any act of man, which causes damages to another, shall oblige the person by whose fault it occurred to repair it’ (cited in Turner 2019: 85, n.18).

⁷⁰⁷ Crotoft (2016: 1386-1396).

ones for which criminal liability is appropriate—but also those ‘unintended but injurious harms’ which are, for Crootof, the proper domain of tort liability.⁷⁰⁸ Indeed, Crootof has argued that ‘[t]he law of armed conflict has a built-in accountability gap’ in that ‘[u]nder international law, there is no individualized remedy for civilians whose property, bodies, or lives are destroyed in war [through] lawful but unintended harmful acts’.⁷⁰⁹ Crootof’s proposal starts from the premise that ‘[c]ivilian harm is endemic to armed conflict’ such that ‘[i]n the absence of a war torts regime, it is astonishing how many foreseeable and preventable harms in armed conflict are deemed lawful accidents’.⁷¹⁰ LAWs, then, ‘expose a gap in the existing legal order’.⁷¹¹

One problem with Crootof’s account is that it does not give legal expression or enforceability to those reparatory moral duties which accrue to participants in enterprises which produced, sold and deployed the LAW in Massacre. Liability, on her view, accrues only to states. Another problem is methodological. Crootof’s approach involves the introduction of a new kind of strict liability mechanism by the creation of a war torts regime. Strict liability exists under the mechanisms of vicarious and product liability and I will go on to argue for the application of these doctrines in cases like Massacre in the next chapter, with some amendments. But to introduce a new kind of strict liability mechanism without first exhausting the resources of our existing strict liability mechanisms seems to me to be conceptually unparsimonious. To the extent that this proposal neither fits nor justifies the existing law as a result, this has important ramifications for the legitimacy of a war torts regime. By contrast, in the next chapter, I articulate and defend some novel reforms to the doctrines of vicarious liability and product liability to ensure appropriate liability attribution in cases like Massacre.

⁷⁰⁸ *Ibid*: 1388.

⁷⁰⁹ Crootof (2022: 1063).

⁷¹⁰ *Ibid*: 1066.

⁷¹¹ Crootof (2016: 1353).

And I will show that the reforms which I propose satisfy the Dworkinian criteria of soundness and fit.

An example of the second kind of approach, namely the adjustment of the fault standard, is Abbott's view that we should introduce a new 'reasonable computer standard' in negligence law.⁷¹² On his view:

'Once a manufacturer establishes that a computer tortfeasor *is* safer than a person, the negligence test should focus on whether the computer's *act* was negligent, rather than whether the computer was negligently designed or marketed. [...] the computer is taking the place of a person in the traditional negligence paradigm, and this paradigm would treat the computer more like a person than a product.'⁷¹³

There are some excellent insights in Abbott's proposal. For one thing, as I discuss further in the next chapter, he identifies the interesting legal territory occupied by autonomous AI systems which may function in the space of legal rights and/or obligations as not (just) products but (also) persons. For another thing, albeit figuratively, Abbott identifies the 'computer' as a 'tortfeasor' which I will argue in the next chapter should be taken quite literally. However, Abbott's proposal falls short in at least three ways. Firstly, his proposal falls short on the Dworkinian criterion of *soundness* because the 'reasonable computer standard' does not express a principle which justifies existing negligence law. On the contrary, Abbott's focus is largely on encouraging the social benefits of automation through the transformation of negligence law.⁷¹⁴ Secondly, Abbott's approach is inadequate from the perspective of *fit* with existing law because it involves imposing negligence liability *principally* on the manufacturer in virtue of the product's failure to meet the reasonable computer standard. In the next chapter,

⁷¹² Abbott (2018).

⁷¹³ *Ibid*: 29.

⁷¹⁴ *Ibid*: 16-24.

I suggest that we should attribute liability *principally* to the LAW in cases like Massacre and *secondarily* to the enterprise which deployed the LAW. My proposal fits with the doctrine of vicarious liability; Abbott's proposal fits neither with that doctrine nor with fault-based negligence liability. Finally, to the extent that Abbott conceives of the 'computer' as a 'tortfeasor', he ought to acknowledge as a corollary of that claim that a tortfeasor—that is, an entity capable of bearing legal responsibility in tort law—must be a legal person. That is an argument which I develop in the next chapter.⁷¹⁵

8.2.2.3. Strict liability

Based on the foregoing discussion, it appears that we need a mechanism of strict liability to give legal expression and enforceability to the reparatory moral duties borne by those who participated in the production, sale and use of the LAW in Massacre. There are two principal mechanisms for the attribution of strict liability in private law: product liability⁷¹⁶ and vicarious liability. This subsection briefly adumbrates why neither mechanism in its current form applies to cases like Massacre. In the next chapter, I offer some principled reforms to amend this 'gap' in legal responsibility.

Strict product liability

Although a comprehensive review of the application of strict product liability frameworks internationally is beyond the scope of this dissertation, in general in order to bring a claim in strict product liability a claimant needs to establish on the balance of probabilities that the

⁷¹⁵ In part, my argument in the next chapter may be construed as a charitable reinterpretation of Abbott's theory: instead of merely *treating* the computer as a person, I submit that we should legally so treat it; I argue for the satisfaction of the traditional conditions for vicarious liability partly on that basis.

⁷¹⁶ There are two alternative kinds of product liability in addition to that strict kind which I focus on here. The first is a form of negligence liability concerning the failure by a manufacturer to fulfil a duty of care in design, production or testing; the second is a form of contractual liability involving the failure of the product to meet express or implied promises of quality or safety. Whilst my earlier arguments regarding the difficulties of applying negligence or contractual liability in cases like Massacre are not decisive, they give sufficient reason to doubt that these alternative forms of product liability would give legal expression or enforceability to those reparatory moral duties which I have defended.

product was defective and unreasonably dangerous at the time it left the defendant's hands, that the product reached the claimant without substantial change and that the defect was the 'proximate cause' of the claimant's injuries.⁷¹⁷ I briefly consider the issues which arise in relation to each of the three limbs of a strict product liability claim in cases like *Massacre*, including a preliminary question as to the legal status of autonomous AI as a product at all.⁷¹⁸ None of these issues, I suggest, is fatal to the application of strict product liability law in cases like *Massacre*; the concerns centre rather on legal uncertainty.

There has been extensive debate as to whether an autonomous AI system is a product or a service. Traditionally, strict product liability applies only to 'tangible goods'.⁷¹⁹ However, AI systems exhibit product-like characteristics:⁷²⁰ they are 'designed, manufactured, sold, and marketed with specific functionalities and promises';⁷²¹ importantly, defects in design, implementation, or deployment cause harm.⁷²² So, in the last two years, the US and EU have introduced legislation classifying AI systems legally as products.⁷²³ This prompts questions concerning the application of the standard three-limb test.⁷²⁴

There is legal uncertainty in the autonomous AI context as to whether unexpected 'decisions', such as that which occurs in *Massacre*, evince 'defects'. As Young notes: 'courts

⁷¹⁷ Second Restatement of Torts, S.402A (US); c.f., Consumer Protection Act 1987 (UK); European Product Liability Directive (85/374/EEC) (EU). Certainly, in the US context a claimant can bring a claim against almost every business in the chain of distribution (*Loomis v. Amazon.com LLC* 63 Cal.App.5th 466 (2021); *Bolger v. Amazon LLC* 53 Cal.App.5th 431 (2020)).

⁷¹⁸ There has been some debate in the EU context as to whether software constitutes a 'product' or a 'service' under product liability legislation; the new EU Product Liability Directive (2024/2853) appropriately establishes software as a product.

⁷¹⁹ Bryan (2024).

⁷²⁰ *Garcia v Character.ai* [2025] Case No. 6:24-CV-01903-ACC-UAM.

⁷²¹ Young (2025); Sharkey (2024).

⁷²² Aligning Incentives for Leadership, Excellence, and Advancement in Development (AI LEAD) Act 2025: S.2937, 119th Cong. §2.

⁷²³ *Ibid*: §1 (US); EU Product Liability Directive (2024/2853).

⁷²⁴ Alberts *et al.* (2025); Smith *et al.* (2024).

and scholars have debated whether defects should be defined by average performance, inherent risk, or user control, especially in autonomous systems.⁷²⁵

In the US, Wu⁷²⁶ has analysed the issue of defectiveness in strict product liability for autonomous vehicles. Courts across the US apply one or more of the following tests:⁷²⁷

- *The ordinary consumer test*: What would an ordinary consumer expect from a product? (This test typically applies where the potential for injury is clear to consumers from the product's nature.)
- *The risk-utility balancing test*: Do the risks associated with the product design outweigh the benefits to the consumer or the public?
- *The prudent manufacturer test*: Would a reasonably prudent manufacturer or seller, aware of the product's dangerous condition, have put the product on the market if it had been aware of the product's condition?
- *The combination test*: Can the manufacturer show a lack of defect? (This test shifts the burden of proof and only applies in exceptional circumstances.⁷²⁸)
- *The ultimate issue test*: The jury may have discretion to determine whether a design is defective.

However, it is unclear whether the LAW in Massacre would be defective on these standard tests. For instance, a leading case in the automotive context is *Anderson v. General Motors*

⁷²⁵ Young (2025); Buiten (2024).

⁷²⁶ Wu (2013; 2016; 2020).

⁷²⁷ Wu (2020: 5-6); Vargo (2014). Although it appears to me rather a form of negligence, it is sometimes called 'strict product liability' where a defendant has failed to provide adequate warnings or instructions to a claimant about the product (Wu 2016: 561-2); under the stipulated idealising assumptions I assume that no such failure to warn occurred in Massacre.

⁷²⁸ C.f. Cane (1993): 355): 'it is usually asserted [that] the onus is on those who wish to shift a loss to justify the shift' given 'administrative expense'.

(GM),⁷²⁹ where a jury found GM liable in the amount of \$4.9bn for burn injuries to six victims due to a defective fuel system in its vehicles. GM had designed vehicles with dangerously positioned fuel tanks which created an unreasonably high risk of fire in low-speed accidents; safer alternatives were feasible but not implemented. But Massacre is disanalogous because there is no apparent design or manufacturing flaw. Rather, the product and design are tried-and-tested, more reliable than a human alternative (e.g., the LAW operates within a reasonable tolerance band) and compliant with the applicable just war principles. As a result, it is unclear whether the LAW would be ‘defective’ under (i)-(iii). Similarly, under (iv), the manufacturer may be able to establish a lack of defect and a jury could conclude that the design was not defective under (v).

In the EU context,⁷³⁰ the consumer-expectation test has traditionally applied, ‘which considers whether a product fails to meet the safety expectations the public is entitled to, considering all relevant circumstances.’⁷³¹ However, the producer ‘may avoid liability if a defect is deemed unavoidable given the scientific and technical knowledge available at the time’.⁷³² Indeed, the Expert Group on Liability and New technologies has alleged that the consumer-expectation test leaves open whether ‘decisions’ by autonomous systems can evince a defect.⁷³³ Whilst this gloss offers no decisive arguments, it suffices to highlight the legal uncertainty surrounding whether the LAW in Massacre is ‘defective’ under standard tests.

In addition, the claimant would need to establish that any defect in the LAW was present at the time that it left the producer’s hands such that the product reached the consumer without substantial change. It has been suggested that this is difficult to establish where a product uses

⁷²⁹ [1985] Civ.21876 Court of Appeal, Third District, California.

⁷³⁰ The UK Consumer Protection Act implemented the European Product Liability Directive (85/374/EEC).

⁷³¹ Buiten (2024: 252); Vansweevelt and Weyts (2009: 509). The next chapter considers reforms to be implemented under the new EU Product Liability Directive (2024/2853).

⁷³² Buiten (2024: 252).

⁷³³ Report from the Expert Group on Liability and New technologies (2019: 28).

autonomous AI: ‘when an AI system continues to learn after deployment [...] [ensuring safety] requires constant testing and high-quality simulators, possibly assisted by AI tools, to maintain and improve their performance.’⁷³⁴ For example, even if a particular feature of the software can be established as defective, it might be challenging in a complex machine learning system for claimants to show that that feature existed at the relevant time and was not a bug ‘learnt’ subsequently.

Finally, some lawyers suggest that it may be difficult for claimants to establish causation of injury due to a particular defect, partly due to what I call ‘technical opacity’—that is, the intrinsic complexity of black-box AI—and partly due to what I call ‘procedural opacity’—that is, opacity due to the difficulties in ensuring disclosure of commercially relevant information, particularly in military technology applications.⁷³⁵ These considerations should prompt us to consider that strict product liability carries promise—at least as a way to give legal expression and enforceability to those reparatory moral duties which accrue to the producers and sellers of LAWs. However, the mechanism needs adapting if it is to ensure justice with adequate certainty following harms involving autonomous AI systems.

Vicarious liability

Vicarious liability renders a defendant D2 (traditionally an employer) liable for the tortious behaviour of another party D1 (traditionally an employee) when acting in the course of his or her employment. Across the common law world, when determining whether D2 should bear vicarious liability for the tortious actions or omissions of D1, the courts apply a two-stage inquiry. At stage one, the court will evaluate the relationship between D1 and D2. Traditionally, the first limb of the test for vicarious liability was the existence of an employment

⁷³⁴ Buiten (2024: 257); Chen *et al.* (2018).

⁷³⁵ Behrendt and Chandler (2025); Norton Rose Fulbright (2024).

relationship.⁷³⁶ However, the law of vicarious liability is ‘on the move’⁷³⁷ and stage one has been broadened to accommodate different types of relationships ‘akin to employment’.⁷³⁸ At stage two, the court will consider whether the wrongful act falls within the scope of the (quasi-)employer’s liability. The traditional test at stage two was the Salmond test which rendered an employer liable for acts ‘which he has not authorized, provided they are so connected with acts which he has authorised that they may rightly be regarded as modes—although improper modes—of doing them.’⁷³⁹ However, as we will see below, in the decision of the Canadian Supreme Court in *Bazley v Curry*,⁷⁴⁰ that test was deemed to give inadequate weight to the policy factors underlying the doctrine of vicarious liability. As such, the courts have adopted a broader ‘close connection’ test, variants of which have been applied in the Supreme Court of Canada;⁷⁴¹ the UK Supreme Court⁷⁴² and the Singapore Court of Appeal.⁷⁴³ This prompts the question: under the doctrine of vicarious liability as it currently stands, could those who produce, sell and use the LAW in cases like Massacre be held liable for its consequences?

There appear to be two principal problems with the application of the doctrine of vicarious liability as it currently stands in cases like Massacre. Firstly, it is unclear whether the standard two-stage test for vicarious liability is met in such cases. At stage one, the relationships between a LAW and those enterprises which produce, sell or use it do not fall within any of those relationships currently recognised as ‘akin to employment’. At stage two, the LAW’s

⁷³⁶ *Honeywill and Stein Ltd v Larkin Brothers Ltd* [1934] 1 KB 191 (CA) 196 (Slesser LJ).

⁷³⁷ *Various Claimants v Catholic Child Welfare Society (CCWS)* [2012] UKSC 56 per Lord Phillips at [19].

⁷³⁸ *JGE v English Province of Our Lady of Charity* at first instance: [2011] EWHC 2871 (QB) (*JGE No. 1*) at [34]–[36] per MacDuff J (affirmed by the Court of Appeal: [2012] EWCA Civ 938 (*JGE No. 2*)).

⁷³⁹ Heuston and Buckley (1996: 443).

⁷⁴⁰ [1999] 2 SCR 534.

⁷⁴¹ *Ibid*; *Jacobi v Griffiths* [1999] 2 SCR 570; *EB v Order of the Oblates of Mary Immaculate in the Province of British Columbia* [2005] 3 SCR 45.

⁷⁴² *Lister v Hesley Hall Ltd* [2001] UKHL 22 (*Lister*); *Dubai Aluminium Co Ltd v Salaam* [2002] UKHL 48 (*Dubai Aluminium*); *CCWS*; *Wm Morrison Supermarkets plc v Various Claimants* [2020] UKSC 12 (*Morrison Supermarkets*).

⁷⁴³ *Skandinaviska Enskilda Banken AB v Asia Pacific Breweries (Singapore) Pte Ltd* [2011] SGCA 22 (*Skandinaviska Enskilda Banken*). It is however worth noting that the Australian jurisprudence has been less enamoured with the *Bazley v Curry* approach (Brodie 2010: ch.1).

‘conduct’ in cases like Massacre would need to satisfy the test for ‘close connection’ with the ‘acts’ which it was authorised to do. In the next chapter I will argue for the introduction of a new category of relationship akin to employment which would accommodate the relationship between a LAW and the military enterprise which *deploys* it and ensure that the LAW’s ‘conduct’ in Massacre satisfies the close connection test. However, the relationship between a LAW and those enterprises which *produce* and *sell* it is not in any sense ‘akin to employment’ so the application of vicarious liability in those relationships falls at the first hurdle. Secondly, vicarious liability is a doctrine of parasitic liability; it requires, therefore, as a corollary of its application in cases like Massacre, that the LAW is capable of bearing liability. But LAWs, lacking legal personhood, do not yet satisfy that necessary precondition for the doctrine’s application.

The application of the doctrine of vicarious liability in cases like Massacre has nowhere been developed at any length. But it is important to note the contribution of Glavaničová and Pascucci.⁷⁴⁴ The authors are concerned principally with the application of liability vicariously to a manufacturer of an autonomous system where an employee has failed to spot a design flaw. They go on to briefly consider the possibility of attributing liability vicariously to the manufacturer for the tortious ‘conduct’ of the autonomous system itself. However, the argument involves three shortcomings which derive from treating vicarious liability as an entirely unprincipled device for accident-prevention and loss-distribution. Firstly, the authors make the *ad hoc* stipulation of a wholly new category of relationship for the application of vicarious liability, namely that which exists between the manufacturer and an autonomous AI system. Secondly, the legal test that the tortious conduct should fall within the scope of employment under the usual doctrine of vicarious liability is replaced on Glavaničová’s and Pascucci’s view by a ‘purpose of manufacturing’ test for deciding the scope of tortious conduct for which

⁷⁴⁴ Glavaničová and Pascucci (2022).

manufacturers might be liable. No principled reason is offered for either of these two fundamental revisions of the standard vicarious liability doctrine and, as I show in the next chapter, they neither fit nor justify the jurisprudence as well as the reforms which I propose. Finally, the authors fail to appreciate that for the doctrine of vicarious liability to apply, the autonomous system would need to be legally personified. It is therefore to the articulation and defence of the first principled argument for the application of vicarious liability in cases like Massacre that I turn in the next chapter.

9. Persons and products

Come, you masters of war, you that build the big guns

You that build the death planes, you that build all the bombs

You that hide behind walls, you that hide behind desks

*I just want you to know I can see through your masks.*⁷⁴⁵

In the last chapter, I argued that those who participate in enterprises which produce, sell and deploy LAWs bear reparatory moral duties to the victims in cases like Massacre which are not adequately reflected in existing private law responsibility mechanisms. This chapter articulates and defends novel proposals for law reform to patch that ‘gap’ in legal responsibility which the previous discussion evinced. In essence, I argue for the modification of the doctrines of vicarious liability and strict product liability to ensure that those who participate in enterprises which pose characteristic risks of harm to others through the production, sale or use of LAWs bear the costs when those risks materialise. Initially, in **9.1-9.2**, I focus on the reparatory moral duties of those participating in military enterprises which *deploy* LAWs. Firstly, in **9.1**, I argue that it would be consistent in principle with common law jurisprudence on the doctrine of vicarious liability to apply that doctrine to cases like Massacre to ensure that militaries bear liability vicariously for the ‘conduct’ of LAWs. However, it would need to be possible for LAWs to be legally responsible for the outcome in Massacre in order for vicarious liability to be derived. In other words, LAWs would need to be legally personified and ascribed a corresponding legal duty. Secondly, in **9.2**, I argue that there is independent reason for LAWs to be instituted as legal persons with a legal duty not to direct attacks against civilians. So, the legal personification of LAWs—and, perhaps, other autonomous AI systems—under international private law would be a novel and timely reform to patch a significant ‘gap’ in

⁷⁴⁵ Dylan (1963).

legal responsibility for autonomous AI.⁷⁴⁶ In **9.3**, I address the legal expression and enforcement of those reparatory moral duties which accrue to participants in enterprises which *produce* and *sell* LAWs, contending that specific modifications to the doctrine of strict product liability would be an appropriate response. I conclude that this two-pronged approach—one which accommodates the reality that autonomous AI systems are functionally both employees and products at private law—would close the only ‘responsibility gap’ which there actually is.

9.1. Vicarious liability at common law

This section argues that the application of the doctrine of vicarious liability in cases like Massacre would satisfy the Dworkinian conditions of *soundness* and *fit* with existing applications of that doctrine across the common law world. In other words, it would satisfy the Dworkinian ‘legislative principle of integrity’ which stipulates that legislation should preserve the moral coherence of the law.⁷⁴⁷ Such an application would satisfy *soundness* given its natural harmony with the enterprise liability rationale for vicarious liability (**9.1.1**). It would also fit with existing applications of that doctrine because, as I argue in the rest of this section, cases like Massacre seem to satisfy the standard two-limb test for the application of the doctrine of vicarious liability. On the first limb, I argue that the ‘relationship’ between a LAW and the military organisation which deploys it should be considered ‘akin to employment’ (**9.1.2**). On the second limb, I argue that the ‘decision’ by the LAW in Massacre is so closely connected with what it is authorised to do that Massacre can fairly and properly be regarded as occurring in the ordinary course of the LAW’s ‘employment’ (**9.1.3**). This prompts the question: can and should the LAW’s ‘conduct’ in Massacre attract liability in the first place? To answer this

⁷⁴⁶ Often, of course, those military enterprises will be part of the executive branch of government. However, military enterprises might also be non-governmental organisations and some governmental organisations with military capacities are not states. So vicarious liability for the military enterprise may in some cases be a form of state liability but that is not necessarily the case.

⁷⁴⁷ See Part One: **4.1.2**; Dworkin (1986: 167, 176).

question, we must first establish the corollary that LAWs can and should be ascribed legal personhood. It is to this task that I turn in **9.2**.

9.1.1. The enterprise liability rationale

As we saw in the last chapter, there has been significant broadening of the doctrine of vicarious liability in recent years, particularly in Canada, the UK and Singapore. This section examines the reasons for this expansion of the doctrine in these three jurisdictions and highlights the particular salience of enterprise liability as a justifying rationale.⁷⁴⁸

9.1.1.1. The Canadian jurisprudence

The justificatory rationale for vicarious liability has historically been controversial. An oft-cited summary is that:

‘Vicarious liability is the creation of many judges who have had different ideas of its justification or social policy, or no idea at all’.⁷⁴⁹

The judgment of the Canadian Supreme Court in *Bazley v Curry* dealt rigorously with that issue, establishing enterprise liability as the fundamental justificatory rationale for vicarious liability in Canada.⁷⁵⁰

‘the employer puts in the community an enterprise which carries with it certain risks. When those risks materialize and cause injury to a member of the public despite the employer’s reasonable efforts, it is fair that the person or organization that creates the enterprise and hence the risk should bear the loss.’⁷⁵¹

⁷⁴⁸ In the Australian context see: Queensland Law Reform Commission (2001: 9-13); *Scott v Davis* (2000) 204 CLR 333; *Hollis v Vabu* (2001) 207 CLR 21. On vicarious liability across the common law world more generally: Giliker (2022).

⁷⁴⁹ Williams (1957: 231).

⁷⁵⁰ Brodie (2007: 494-495); Giliker (2016: 161). See also: *John Doe v Bennett* [2004] 1 SCR 436. For criticism of the Canadian jurisprudence: McCarthy (2004).

⁷⁵¹ *Bazley v Curry* at 60.

This reflects a widely held view in the legal philosophy literature that, per Fleming:

‘[m]ost important of these [justificatory rationales for vicarious liability] is the belief that a person who employs others to advance his own economic interest should in fairness be placed under a corresponding liability for losses incurred in the course of the enterprise’.⁷⁵²

Indeed, Atiyah has it that:

‘The master ought to be liable for all those torts which can fairly be regarded as reasonably incidental risks to the type of business he carries on.’⁷⁵³

The moral nerve, as we saw in the last chapter, appears to be some concept of fairness. What is the nature and scope of the associated legal duties?

As to their nature, the sister case before the Canadian Supreme Court of *Jacobi v Griffiths* indicates that vicarious liability reflects backwards-looking duties of corrective justice. McLachlin J (on behalf of L’Heureux-Dubé, McLachlin and Bastarache JJ) stated in the dissenting judgment that:

‘the key question is whether the Club’s operation of its business and use of Griffiths in his employment position created or materially enhanced the risk of the sexual assaults that took place [...] [such that the Club] should be held financially responsible for those risks when they materialized’.⁷⁵⁴

Binnie J, giving the majority judgment of the Supreme Court in *Jacobi v Griffiths*, agreed on that ‘key question’ but found that vicarious liability was not engaged on the facts since the enterprise had not materially increased risk of harm.⁷⁵⁵ Setting aside that factual disagreement,

⁷⁵² Fleming (1998: 410). See also: Atiyah (1967: 24, 171); Cane (1993: 85).

⁷⁵³ Atiyah (1967: 171).

⁷⁵⁴ *Jacobi v Griffiths* at [12], [16].

⁷⁵⁵ *Ibid* at [79]-[86].

the test of a material increase in risk also specifies the scope of vicarious liability, limiting the reparatory duties engaged to the extent of the material increase in risk.

9.1.1.2. The UK context⁷⁵⁶

The Canadian jurisprudence has had a significant impact in the UK context. In *Lister v Hesley Hall Ltd*, the House of Lords reversed previous English authority,⁷⁵⁷ holding the owners and managers of a school vicariously liable for sexual assaults committed by the warden of a boarding house. Lord Steyn referred to *Bazley v Curry* and *Jacobi v Griffiths* as ‘luminous and illuminating’⁷⁵⁸ judgments which would henceforth be the ‘starting point’ for consideration of similar cases across the common law world.⁷⁵⁹ Lord Millett addressed the underlying policy considerations, citing enterprise liability as the most plausible basis for vicarious liability⁷⁶⁰ and concluding that:

‘[i]f the employer’s objectives cannot be achieved without a serious risk of the employee committing the kind of wrong which he has in fact committed, the employer ought to be liable [...] if the risk is one which experience shows is inherent in the nature of the business.’⁷⁶¹

Similarly, in *Dubai Aluminium Co Ltd v Salaam*, Lord Nicholls gave the majority judgment of the House of Lords:⁷⁶²

‘The underlying legal policy [for vicarious liability] is based on the recognition that carrying on a business enterprise necessarily involves risk to others. It involves the risk

⁷⁵⁶ Judgments of the House of Lords and the Supreme Court may apply to Scotland and Northern Ireland too but otherwise the jurisprudence cited is binding only in England and Wales.

⁷⁵⁷ *Trotman v North Yorkshire County Council* [1999] LGR 584.

⁷⁵⁸ *Lister* at [27].

⁷⁵⁹ *Ibid.*

⁷⁶⁰ *Ibid* at 243: G-H.

⁷⁶¹ *Ibid* at 244: A-B.

⁷⁶² *Dubai Aluminium* at 21.

that others will be harmed by wrongful acts committed by the agents through whom the business is carried on. When those risks ripen into loss, it is just that the business should be responsible for compensating the person who has been wronged.’

Subsequently, the Court of Appeal imposed vicarious liability on an employer for an employee’s commission of harassment on the basis that harassment is a ‘risk incidental to employment’.⁷⁶³

In the Catholic Child Welfare Society case (*CCWS*), Lord Phillips (giving the judgment of the Supreme Court) identified five factors which ‘usually make it fair, just and reasonable to impose vicarious liability’ in relationships akin to employment.⁷⁶⁴ These are: (i) ‘[t]he employer is more likely to have the means to compensate the victim than the employee and can be expected to have insured against that liability’; (ii) ‘[t]he tort will have been committed as a result of activity being taken by the employee on behalf of the employer’; (iii) ‘[t]he employee’s activity is likely to be part of the business activity of the employer’; (iv) ‘[t]he employer, by employing the employee to carry on the activity will have created the risk of the tort committed by the employee’; and (v) ‘[t]he employee will, to a greater or lesser degree, have been under the control of the employer’.⁷⁶⁵ We will consider the interpretation of the *CCWS* test in greater detail below but it has been upheld in consistent judgments by the Supreme Court that (i) and (v) have limited significance.⁷⁶⁶ As such, the *CCWS* test for a relationship capable of giving rise to vicarious liability is that:

‘(1) the tort will have been committed as a result of activity being taken by the tortfeasor on behalf of the defendant, (2) the tortfeasor’s activity is likely to be part of the business

⁷⁶³ *Majrowski v Guy’s & St Thomas’s NHS Trust* [2005] IRLR 340 at 350. Note that the Court of Appeal’s decision was upheld on appeal on a different basis.

⁷⁶⁴ *CCWS* at [35].

⁷⁶⁵ *Ibid.*

⁷⁶⁶ *Cox v Ministry of Justice* [2016] UKSC 10 (*Cox*) at [20]-[21]; *Armes v Nottinghamshire County Council* [2017] UKSC 60 (*Armes*) at [56].

activity of the defendant, and (3) the defendant, by employing the tortfeasor to carry on the activity, will have created the risk of the tort committed by the tortfeasor.⁷⁶⁷

Although as I discuss below this test is not sufficient for vicarious liability,⁷⁶⁸ it affirms in my submission the central justificatory role of enterprise liability in the English doctrine of vicarious liability.⁷⁶⁹

9.1.1.3. The Singaporean jurisprudence

In 2011, the Singapore Court of Appeal in *Skandinaviska Enskilda Banken* endorsed the approach taken by the Supreme Court of Canada in *Roman Catholic Episcopal Corporation of St George's v John Doe and another*⁷⁷⁰ where McLachlin CJ identified the primary rationale of vicarious liability as enterprise liability:

‘Vicarious liability is based on the rationale that the person who puts a risky enterprise into the community may fairly be held responsible when those risks emerge and cause loss or injury to members of the public. Effective compensation is a goal. Deterrence is also a consideration. The hope is that holding the employer or principal liable will encourage such persons to take steps to reduce the risk of harm in the future.’⁷⁷¹

As Tan discusses,⁷⁷² the enterprise liability rationale for vicarious liability was subsequently affirmed in a rare five-member full bench decision by the Singapore Court of Appeal in *Ng Huat Seng v Munib Mohammad Madni*.⁷⁷³ Menon CJ (delivering the unanimous judgment of

⁷⁶⁷ *Armes* at [57]; *Cox* at [24].

⁷⁶⁸ *Barclays Bank v Various Claimants* [2020] UKSC 13 (*Barclays Bank*) at [16] per Lady Hale; *Lister* at [60] per Lord Hobhouse.

⁷⁶⁹ I suggest that enterprise liability also explains the longstanding distinction in the vicarious liability jurisprudence between employees and independent contractors. Vicarious liability does not extend to independent contractors because ‘whether the tortfeasor is carrying on business on his own account’ (*Barclays Bank* at [24], [27])—that is, ‘taking the risk of profit or loss’ (Giliker (2022: 78); *JGE No.2* at [70] per Ward LJ)—is a decisive factor.

⁷⁷⁰ [2004] 1 SCR 436.

⁷⁷¹ *Ibid* at [20], cited in *Skandinaviska Enskilda Banken* at [70].

⁷⁷² Tan (2022).

⁷⁷³ [2017] SGCA 58, [2017] 2 SLR 1074.

the Court) considered the recent English and Welsh case law on vicarious liability and affirmed the application of the law as stated by Lord Phillips in *CCWS* in Singapore law.⁷⁷⁴ Concurring with Lord Phillips' statement that the mere creation of risk was not, of itself, enough to give rise to vicarious liability, Menon CJ highlighted that in addition to the special relationship between D1 and D2:

‘the defendant must in some way have *created or significantly enhanced, by virtue of that relationship, the very risk that in fact materialised* in order to be held vicariously liable for the tortfeasor’s wrongful acts’.⁷⁷⁵

So, in Singapore too, enterprise liability is recognised as a core justificatory rationale for vicarious liability.

The arguments advanced in this section make no claim to be decisive. But I believe that they give us sufficient reason to think that a central justificatory rationale for the application of vicarious liability in any given factual scenario is enterprise liability. As we saw in the last chapter, that same justificatory rationale imposes reparatory moral duties on those who participate in the enterprise(s) which deployed the LAW in Massacre.⁷⁷⁶ I submit that the legal expression and enforcement of those reparatory moral duties via the mechanism of vicarious liability would therefore satisfy Dworkinian *soundness*. Let us turn, then, to the question of whether such an application of vicarious liability would also *fit* with the existing case law.

9.1.2. Beyond employment: relationships reinterpreted

This subsection argues for the introduction of a new category of special relationship at stage one of the standard two-limb test for vicarious liability. This category would be better attuned than the traditional ‘employment test’ to the reality of an automated era in which enterprises

⁷⁷⁴ *Ibid* at [61]-[66].

⁷⁷⁵ *Ibid* at [66].

⁷⁷⁶ As well as on those who participate in enterprises which produced and sold the LAW, as I discuss in 9.3.

are able to create risks to others in new ways. In brief: not just through the *people* who they employ as agents but also through *autonomous AI* ‘agents’ which they deploy, often instead of humans. Where the risks associated with a particular autonomous AI system are especially severe—through, say, a small chance of a very severe harm—the need to legally enforce the moral duty on enterprises to internalise their own risks becomes especially pressing. In brief, the core thread of my argument in what follows is that in cases like Massacre the imposition of vicarious liability would fit with the case law on the identification of a special relationship between the enterprise and the ‘tortfeasor’ at stage one. As we have seen, there is broad agreement on the nature of this test across the common law world so it will suffice for my purposes to establish that the identification of the ‘relationship’ between a LAW and the military enterprise which deploys it as ‘akin to employment’ would fit with English and Welsh jurisprudence on stage one of the standard two-limb test for vicarious liability. This gives good reason to suppose that there would also be fit with wider common law jurisprudence.

Firstly, in addition to the traditional employment relationship, it is now well established that other categories of relationships may satisfy the stage one test for vicarious liability provided that they are ‘akin’ to employment. Relationships ‘akin’ to employment include, but are not limited to: the relationship the bishop and a priest in a diocese;⁷⁷⁷ the relationship between teachers in a school and their associated religious organisation;⁷⁷⁸ the relationship between the Ministry of Justice and prisoners serving in the prison kitchen for rehabilitation;⁷⁷⁹ and the relationship between foster parents volunteering to undertake the care of children placed under local authority control and that local authority.⁷⁸⁰ My proposal is to introduce a

⁷⁷⁷ *JGE No.1* at [34]–[36] per MacDuff J (affirmed by the Court of Appeal in *JGE No.2*); c.f. the Canadian case of *John Doe v Bennett* at [27] per McLachlin CJ.

⁷⁷⁸ *CCWS* at [49]. The Court found there to be joint vicarious liability between the religious organisation and the school for the tortious actions of the teachers.

⁷⁷⁹ *Cox*.

⁷⁸⁰ *Armes*.

new category, namely the relationship between the military enterprise which deploys the LAW and the LAW itself. We will see below why there is independent reason to allow the LAW itself to be the principal tortfeasor at law. But in the rest of this subsection, I explain why the application of vicarious liability to the new category of relationship which I have proposed would be consistent with the stage one test as set out in *CCWS* and upheld in *Cox*, *Armes* and *Barclays Bank*.

In *Cox*, the Supreme Court analysed the features which Lord Phillips identified in *CCWS* as indicators that a relationship is sufficiently akin to employment that it is usually fair, just and reasonable to impose vicarious liability. Lord Reed JSC (with whom Lord Neuberger PSC, Baroness Hale DPSC, Lord Dyson MR and Lord Toulson TSC agreed) concluded that Lord Phillips had developed ‘a modern theory of vicarious liability’⁷⁸¹ according to which:

‘a relationship other than one of employment is in principle capable of giving rise to vicarious liability where harm is wrongfully done by an individual who carries on activities as an integral part of the business activities carried on by a defendant and for its benefit (rather than his activities being entirely attributable to the conduct of a recognisably independent business of his own or of a third party), and where the commission of the wrongful act is a risk created by the defendant by assigning those activities to the individual in question.’⁷⁸²

The Court also emphasised that a ‘narrow’ reading of ‘business’, ‘benefit’ and ‘enterprise’ is misleading, noting that there is a ‘long established application of vicarious liability to public authorities and hospitals’ and that ‘the benefit which [the enterprise] derives from the tortfeasor’s activities [need not] take the form of a profit’.⁷⁸³

⁷⁸¹ *Cox* at [24].

⁷⁸² *Ibid.*

⁷⁸³ *Ibid* at [30].

Applying the test set out in *CCWS*, the Court in *Cox* held that a prisoner working in the prison kitchen as part of his rehabilitation satisfied the stage one (and stage two) test(s) for vicarious liability:

‘[activities assigned to] [p]risoners working in the prison kitchens [...] form an integral part of the activities which [the prison service] carries on in the furtherance of its aims [...]. They are placed by the prison service in a position where there is a risk that they may commit a variety of negligent acts within the field of activities assigned to them. [...] Furthermore, they work under the direction of prison staff. Mrs Cox was injured as a result of negligence by Mr Inder in carrying on the activities assigned to him. The prison service is therefore vicariously liable to her.’⁷⁸⁴

I suggest that we could draw analogous conclusions regarding the relationship between the LAW in Massacre and the military organisation which deployed it. The deployment of the LAW is an integral part of the activities which the military organisation carries on in furtherance of its aims, aims which include the deployment of safe and effective technology. LAWs are placed by the military ‘in a position where there is a risk that they may commit a variety of [‘]negligent acts[’] within the field of activities assigned to them’. And LAWs work under the direction of military officers. The victims in a case like Massacre are injured or killed as a result of something akin to negligence by the LAW, as I discuss in **9.2**, in carrying out the activities assigned to it. The military organisation is therefore vicariously liable to the victims.

The Supreme Court applied *CCWS* and *Cox* in the subsequent case of *Armes*. That case concerned whether the local authority could be held vicariously liable for the abuse of a child by foster parents with whom she was placed while in its care. Lord Reed JSC (with whom Baroness Hale, Lord Kerr and Lord Clarke agreed, Lord Hughes JSC dissenting) gave the

⁷⁸⁴ *Ibid* at [32].

judgment of the Supreme Court that the relationship between the local authority and the foster parents satisfies the stage one (and stage two) test(s) for vicarious liability. An interesting point for the application of vicarious liability to autonomous AI emerges at this juncture. The Supreme Court has consistently held that in identifying a relationship capable of incurring vicarious liability ‘the significance of control is that the defendant can direct what the tortfeasor does, not how he does it’.⁷⁸⁵ However, the reflections of the majority in *Armes* on the relationship between control and risk creation are particularly instructive in cases like Massacre:

‘the local authority[...] [places] children in their care with foster parents [...] in circumstances where close control cannot be exercised by the local authority [...]. Although it is generally considered to be in the best interests of children in care that they should be placed in foster care [...] it is relevant to the imposition of vicarious liability that a particular risk of abuse is inherent in that choice. That is because, if the public bodies responsible for decision-making in relation to children in care consider it advantageous to place them in foster care, notwithstanding the inherent risk that some children may be abused, it may be considered fair that they should compensate the unfortunate children for whom that risk materialises [...] In that way, the burden of a risk borne in the general interest is shared, rather than being borne solely by the victims.’⁷⁸⁶

Once again, I suggest that we could say something similar regarding the balancing of the general good with particular risks in cases like Massacre. Let us rephrase the passage above for cases like Massacre. The military deploys LAWs in circumstances where close control cannot be exercised by the military. Although it is generally considered in such cases to be in the best

⁷⁸⁵ *CCWS* at [36]; *Cox* at [21]; *Armes* at [56].

⁷⁸⁶ *Armes* at [61].

interests of non-combatants that LAWs are deployed, given their presumed superiority in distinguishing legitimate from illegitimate targets, a particular risk of unlawful engagement is inherent in that choice. This is relevant to the imposition of vicarious liability: if the military considers it advantageous to deploy LAWs, notwithstanding the inherent risk that some non-combatants may be unlawfully targeted, it may be considered fair that the military should compensate the unfortunate non-combatants for whom that risk materialises.

This is not to say that the policy factors set out in *CCWS* and refined in *Cox* and *Armes* are necessary or sufficient for satisfying the stage one test for vicarious liability. Indeed, Lady Hale in *Barclays Bank* endorsed the *dictum* of Lord Hobhouse in *Lister* that

‘an exposition of the policy reasons for a rule [...] is not the same as defining the criteria for its application.’⁷⁸⁷

However, the Court acknowledged:

‘In doubtful cases, the five ‘incidents’ identified by Lord Phillips may be helpful in identifying a relationship which is sufficiently analogous to employment to make it fair, just and reasonable to impose vicarious liability.’⁷⁸⁸

So, based on the arguments advanced in this section, I submit that it would fit with the existing law on vicarious liability to recognise that those organisations which deploy LAWs stand in such a relationship to LAWs as to satisfy the stage one test.

9.1.3. Close connections

The stage two test for vicarious liability is whether the primary tortfeasor’s wrongful conduct was ‘so closely connected with acts which [s]he was authorised to do that [...] it can fairly and properly be regarded as done by [her] while acting in the ordinary course of [her]

⁷⁸⁷ *Lister* at [60]; *Barclays Bank* at [16] per Lady Hale.

⁷⁸⁸ *Barclays Bank* at [27].

employment.⁷⁸⁹ For example, in *Lister*, the House of Lords held that the warden's employment included close contact with pupils and, therefore, certain inherent risks of abuse such that the acts of abuse which the warden committed 'were so closely connected with his employment that it would be fair and just to hold the employers vicariously liable'.⁷⁹⁰ Similarly, in *Dubai Aluminium* the House of Lords found that the wrongful conduct of Mr Amhurst in drafting fraudulent consultancy agreements was 'within the ordinary course of the firm's business [...] even though they were drafted for a dishonest purpose'.⁷⁹¹ By contrast, in *Morrison Supermarkets*, the Supreme Court held that the wrongful conduct of Mr Skelton, a grudge-bearing employee who shared the private information of a number of other employees of Morrison Supermarkets via his personal USB stick on his home computer, was not sufficiently closely connected with acts which he was authorised to do.⁷⁹²

There is another, slightly broader formulation of the stage two test which the Supreme Court applied in *Mohamud*.⁷⁹³ In that case, Lord Toulson stated that:

'the court has to consider two matters. The first question is what functions or 'field of activities' have been entrusted by the employer to the employee [...] Secondly, the court must decide whether there was sufficient connection between the position in which he was employed and his wrongful conduct to make it right for the employer to be held liable under the principle of social justice.'⁷⁹⁴

Under this formulation of the stage two test, wrongful conduct exhibiting 'sufficient connection' to the field of activities entrusted to the employee has included: a racist attack on

⁷⁸⁹ *Morrison Supermarkets* at [46] (Lord Reed); *Dubai Aluminium* at [23] (Lord Nicholls).

⁷⁹⁰ *Lister* at [28] (Lord Steyn).

⁷⁹¹ *Dubai Aluminium* at [36] (Lord Nicholls).

⁷⁹² *Morrison Supermarkets* at [46] (Lord Reed).

⁷⁹³ *Mohamud v Morrison Supermarkets plc* [2016] UKSC 11 (*Mohamud*).

⁷⁹⁴ *Ibid* at [44]-[45] (Lord Toulson).

a customer who had inquired about printing facilities at a supermarket kiosk;⁷⁹⁵ a drunken brawl between the managing director and a colleague during a post-Christmas party at a local hotel;⁷⁹⁶ and a fight and chase around a building site culminating in one labourer striking the other with a scaffolding pole.⁷⁹⁷

The broader formulation of the stage two test has been criticised⁷⁹⁸ and arguably superseded.⁷⁹⁹ However, I submit that it would fit the case law to find that the ‘conduct’ of the LAW in cases like *Massacre* satisfies the stage two test on its narrow or broad formulations. Firstly, the narrow formulation. As in *Lister* and *Dubai Aluminium* and unlike *Morrison Supermarkets*, the LAW’s wrongful ‘conduct’ in *Massacre* occurs within the ordinary course of the LAW’s deployment, even though it is an unauthorised mode of doing authorised activities. Secondly, the broader formulation. The field of activities entrusted to the LAW includes selection and engagement of targets without human intervention. There seems to be sufficient connection between that field of activities and the selection and engagement of unlawful targets that it would be just to hold the military vicariously liable.

These considerations prompt the following objection, foreshadowed in the last chapter. There is a significant disanalogy between cases like *Massacre* and the vicarious liability jurisprudence: the principal ‘wrongdoer’ in *Massacre* is not yet capable of bearing liability. As we saw in the last chapter, a corollary for the application of vicarious liability is that there is a principal tortfeasor. That in turn requires the legal personhood of the principal tortfeasor. In the next subsection, I prove that corollary by showing that we have independent reason to legally personify the LAW and attribute liability in cases like *Massacre* to the LAW itself.

⁷⁹⁵ *Mohamud*.

⁷⁹⁶ *Bellman v Northampton Recruitment Ltd* [2018] EWCA Civ 2214.

⁷⁹⁷ *Levitt v Euro Building and Maintenance Contractors Ltd* [2019] EWHC 2926 (QB).

⁷⁹⁸ Plunkett (2016: 561).

⁷⁹⁹ *Morrison Supermarkets*.

9.2. Legal personhood and liability for LAWs

Recall the argument set out in Part One, 4.1.3. There, I articulated and defended a functionalist conception of legal personhood according to which the ascription of legal personhood to some entity does not depend upon the intrinsic moral status or value of that entity but rather upon its effective performance in the space of moral and legal obligations. This section assumes that same functionalist conception of legal personhood (9.2.1) and argues that LAWs should be ascribed legal personhood (9.2.2), including a legal duty not to target civilians (9.2.3).

9.2.1. Legal personhood: a brief recap

Recall the account which I defended in Part One. A legal person is a moral person which is able to perform effectively in the space of legal obligations. A moral person, applying List and Pettit's account, is 'an agent who can perform effectively in the space of moral obligations'.⁸⁰⁰

There are three conditions for agency according to the functionalist conception:

For some entity X and some context C, X is an agent in C iff:⁸⁰¹

A1.X has representational states that depict how things are in the environment in C;

A2.X has motivational states that specify how X requires things to be in that environment;

and

A3.X has the capacity to process its representational and motivational states, leading it to intervene suitably in the environment in C whenever that environment fails to match a motivating specification.

In Part One, we saw how certain groups, namely 'corporations', are agents in certain contexts on List and Pettit's account. I argued that SARs were also agents in certain contexts on two bases. Firstly, I argued that SARs are likely to pass a suitably specified imitation game such

⁸⁰⁰ List and Pettit (2011: 173).

⁸⁰¹ *Ibid*: 19-20.

that there is good reason to infer that they would pass Dennett's intentional stance test and, as a result, satisfy (A1)-(A3).⁸⁰² Secondly, I argued that group agents may be conceptualised as special cases of AI systems, where the 'hardware' supporting their AI is social rather than electronic or robotic. In **9.2.2**, I submit that LAWs are agents in certain contexts on similar bases.

In addition to this, I argued that SARs were moral persons, that is, agents which perform effectively in the space of moral obligations. According to List and Pettit, to perform effectively in the space of moral obligations, an agent must satisfy a further three conditions, which are necessary and sufficient for the agent to 'contract[...] obligations to others and [...] derive[...] entitlements from the reciprocal obligations of others'.⁸⁰³ For some agent X in C, X is a moral person in C iff:⁸⁰⁴

A4.X faces a normatively significant choice in C;

A5.X has the understanding and access to evidence required for making judgments about the options in C;

and

A6.X has the control required to make decisions in C.

In **9.2.2** I articulate some reasons to believe that LAWs would also satisfy (A4)-(A6) and are, as such, appropriate candidates for legal personhood. Then, in **9.2.3**, I show that as quasi-employees of the military organisation which deploys them, LAWs should bear appropriate legal duties.

⁸⁰² Part One: **4.2.1**.

⁸⁰³ List and Pettit (2011: 173).

⁸⁰⁴ *Ibid*: 155, ch.8.

9.2.2. Legal personhood for LAWs

Firstly, I submit that the LAWs used in cases like Massacre must as a matter of logical necessity satisfy a suitably specified Turing test which, as per the arguments advanced in Part One, implies that the LAWs will satisfy Dennett's intentional stance test. This is good evidence that the systems will satisfy (A1)-(A3). Consider the following reasoning to see why. A suitably specified Turing test for LAWs would require more sophisticated capabilities than Turing's original test. Indeed, as we saw in chapter one, in cases like Massacre, the LAWs will need to perform at least as well or better than a human alternative in the accurate selection and engagement of targets in accordance with the requirements of *jus in bello* and international humanitarian law. I suggest that this gives us an indication of a suitable specification for a Turing test: the LAWs used in cases like Massacre should perform at least as well or better than a human combatant on the battlefield.⁸⁰⁵ So I suggest that the LAWs used in Massacre will by definition satisfy a suitably specified Turing test, which is good evidence that they will satisfy (A1)-(A3).

One might challenge the inference from satisfaction of a suitably specified Turing test to the satisfaction of the functionalist agency conditions. I do not purport to offer any decisive arguments. But an illustration may assist. I have stated that the LAW in cases like Massacre must perform at least as well or better than a human alternative in the accurate selection and engagement of targets in accordance with the requirements of *jus in bello* and international humanitarian law. Clearly, in order to achieve this level of functionality, the LAW must have representational states that depict how things are in the environment (A1)—for instance, to detect the presence of enemy combatants in the environment; motivational states that specify how the LAW requires things to be in that environment (A2)—for instance, the LAW may have

⁸⁰⁵ Suppose, to the contrary, that the LAW in Massacre did not satisfy such a Turing test. In that case, it should not have been used at all and the military organisation which deployed a defective LAW would bear criminal liability for intentionally or recklessly causing unjustified lethal harm. So, the 'gap' in legal responsibility would not arise.

been deployed to ensure that the environment is safe for friendly combatants to enter; and the capacity to process its representational and motivational states, leading it to intervene suitably in the environment in C whenever that environment fails to match a motivating specification (A3)—so, the LAW will recognise those enemy combatants as a target which it is necessary to select and engage in order to ensure a safe environment for friendly combatants. LAWs, then, are functionally agents.

In addition, LAWs appear to be *moral* agents, that is, agents which perform effectively in the space of moral obligations. To illustrate, consider a LAW used in a military context. Because I have argued that we may take the intentional stance towards the LAW, I will use phrases such as ‘the LAW understands/judges/determines’. Suppose that the LAW is deployed in a context where there is an unarmed civilian. Since the LAW is autonomous, it faces the choice as to how to respond: should it engage the target, pass by peacefully or report the findings back to its superior officer? Any one of these ‘choices’ is morally significant. As a result, the LAW satisfies (A4). The LAW understands that the unarmed civilian is not an immediate threat and has sufficient evidence from its training data to judge that engaging the target is therefore not an appropriate option, that to pass by peacefully is necessary but that such findings must also be reported back to its superior officer. As such, the LAW seems to satisfy (A5). Based upon its internal evaluation of what would best satisfy its mission objectives, the LAW determines its response, perhaps passing by peacefully and reporting the findings back to its superior officer (A6). I submit that LAWs, then, seem to be moral agents according to the functionalist conception of moral agency. According to the functionalist conception of legal personhood which I defended in Part One, LAWs are also, therefore, candidates for legal personhood. That is, LAWs would be appropriate holders of legal rights or bearers of legal duties provided that there is at least one legal right or duty which may

justifiably be ascribed to them. In the next subsection, I turn to the consideration of which kinds of legal duties LAWs may appropriately bear.

9.2.3. Liability for LAWs

Recall from Part One the distinction between intrinsic and extrinsic moral status. I argued there that according to the functionalist conception which I have defended, the ascription of legal personhood to some entity does not depend upon the intrinsic moral status or value of that entity. Entities, such as corporations, which lack intrinsic moral status may therefore be legal persons in virtue only of their extrinsic moral status. The distinction between intrinsic and extrinsic moral status sustains a second distinction between non-derivative legal rights, which a legal person has in virtue of her intrinsic moral status, and derivative legal rights, which a legal person has in virtue of the kinds of values or goods that they promote.

I submit that an analogous distinction should be drawn between what I call non-derivative legal responsibility and derivative legal responsibility. Non-derivative legal responsibility is, in essence, liability which is ascribed to a legal person in virtue at least in part of her moral responsibility. Much of criminal liability is like this. I make no claim here about the *sufficiency* of moral responsibility for a bad outcome for the ascription of criminal liability for that outcome. But it is usually necessary for the ascription of criminal liability to some agent in respect of a bad outcome that she also bears moral responsibility for that outcome. Derivative legal responsibility, by contrast, is that kind of liability which a legal person bears in virtue of the kinds of values or good which such an ascription of liability promotes. Ascriptions of liability, in criminal or in private law, to entities like corporations are typically justified by appeal to the desideratum of an efficient system of economic exchange. Vicarious liability in private law is also attributed not in virtue of an agent's moral responsibility but rather, as per the enterprise liability rationale which we explored in the last section, to promote values such as fairness or non-exploitation and policy desiderata including accident-prevention

and loss-distribution. In this subsection, I argue for the ascription of a derivative legal duty to LAWs not to direct attacks against civilian populations, individual civilians or civilian objects.

The central valuable good upon which I wish to focus is the good of giving legal expression and enforceability to the moral rights of victims and their families to significant reparations in cases like Massacre. Indeed, we saw in the last chapter that enterprise liability seems to justify significant reparatory moral duties which are owed to victims and their families in cases like Massacre as part of the internalisation of their characteristic risks by the enterprises involved in the production, sale and use of the LAW. I argued that one way to give legal expression and enforceability to those reparatory moral duties which are owed by military enterprises involved in the deployment of the LAW would be via the mechanism of vicarious liability. The only catch is that vicarious liability depends upon there being wrongful conduct by the principal. In cases like Massacre, the principal is the LAW and in the eyes of the law it is not capable of ‘wrongful conduct’. I suggest that this is unprincipled and generates counterintuitive consequences.

Firstly, it is unprincipled: the LAW in Massacre as per my arguments in this chapter is in a relationship akin to employment with the military enterprise which deploys it. Functionally, the LAW performs as an employee. Employees of military enterprises are under a legal duty not to intentionally direct attacks against civilian populations, individual civilians or civilian objects.⁸⁰⁶ When that duty is breached, full reparations are owed to the victims and their families.⁸⁰⁷ However, under the stipulated idealising assumptions in cases like Massacre, where the LAW directs an attack on civilian populations, individual civilians or civilian objects, no

⁸⁰⁶ Rome Statute, Art.8(b)(i)-(ii).

⁸⁰⁷ Draft Articles on the Responsibility of States for Internationally Wrongful Acts (‘Draft Articles’), Art.31. The Draft Articles assist in identifying customary international law. C.f. ICRC Rule 156. For discussion: Chengeta (2016).

legal duty is breached.⁸⁰⁸ As a result, no reparations are owed.⁸⁰⁹ This seems to draw a legal distinction without a moral difference. From the perspective of enterprise liability, reparations to the civilians are owed in both cases.

One might object that there is a moral difference because in the case of a human perpetrator that employee was blameworthy, whereas in cases like Massacre no one is. So, the failure of international law to ensure reparations in the latter case reflects a pattern of blameworthiness. Indeed, we may recall the case of Responsible Soldier where, as set out in **6.3.2**, a case analogous to Massacre occurs by the free and knowing conduct of a human soldier. In that case, reparations would be owed because, assuming that there was no justification or excuse for the soldier's conduct, the soldier is blameworthy with respect to the outcome. But in Psychotic Soldier—a case analogous to Massacre except that the harm is perpetrated by a psychotic soldier—reparations would not be owed given the operation of an insanity defence under international criminal law⁸¹⁰ which reflects the non-blameworthy conduct of the psychotic soldier. So, the objection goes, international law appropriately reflects the non-blameworthiness of all parties in Massacre.

The objection mistakes the nature of the problem. The law currently allows LAWs to substitute into the roles of human employees of the military with none of the corresponding legal duties. By contrast, human employees bear legal duties and may subsequently benefit from excuses for their conduct. Because there is no legal duty breached in cases like Massacre, international law currently fails to give expression to the fact that the legal rights of non-combatants are engaged at all. In effect, cases like Massacre are automatically excused and consequently unrepaired.

⁸⁰⁸ Rome Statute, Art.25; UN 2019.

⁸⁰⁹ Draft Articles, Art.3. In particular, the second limb of the Art.3 test—namely that there has been a breach of an international legal duty—is not currently satisfied in Massacre.

⁸¹⁰ Rome Statute, Art.31(1)(a).

This is counter-intuitive because those reasons for excusing morally problematic conduct in human combatants do not apply to the use of LAWs. This is the ‘problem of permissiveness’ which Eggert identifies,⁸¹¹ viz., the problem that applying the same principles to regulate LAWs and human combatants in warfare would be excessively permissive in relation to LAWs. This is because LAWs, unlike human persons, are ‘not subject to moral powers, they cannot be thought to be blameworthy, morally responsible, or excused for harming people’.⁸¹² Eggert argues that, at least in principle, ‘human combatants may be absolved from blame for harming morally innocent individuals, whereas nothing of the sort is true for robots’.⁸¹³ I suggest that the failure to give legal expression to that difference prevents international law from giving adequate legal expression to the moral rights of non-combatants to significant reparations in cases like Massacre.

My proposal is that the ascription of a legal duty to LAWs not to direct attacks against civilian populations, individual civilians or civilian objects would assist in the solution of this problem. Because it contains no mental element, that duty gives legal expression to the fact that its breach by a LAW would be neither blameworthy nor excusable. This legal duty assists in solving the problem in cases like Massacre in at least two ways, one which concerns the legal enforcement of moral reasons and the other which concerns their legal expression. Firstly, this duty would open the door to the ascription of vicarious liability in cases like Massacre to the military enterprise which deployed the LAW. In principle, that would enable the legal enforcement of its significant reparatory moral duties. For another thing, as we saw in Part One, there is an educative or expressive function of the law which means that laws may ‘help to reconstruct norms and the social meaning of action’.⁸¹⁴ I suggest that the ascription of a legal

⁸¹¹ Eggert (2023: 11-12). Eggert’s discussion of narrow alignment also considers related problems in relation to autonomous vehicles which fall outside the scope of my discussion here.

⁸¹² *Ibid*: 12.

⁸¹³ *Ibid*. See also: McMahan (2009).

⁸¹⁴ Sunstein (1996: 2032).

duty to LAWs not to direct attacks against civilian populations, individuals or objects reconstructs cases like Massacre not as accidents but as violations. That may help to construct norms of military engagement and the social meaning of action in ways which better protect the moral rights of non-combatants.

Two closing thoughts on the legal personhood of LAWs. Firstly, lots of authors have been concerned that the ascription of legal personhood to certain kinds of autonomous systems might act as some kind of ‘liability shield’, enabling organisations which deploy autonomous AI systems to avoid liability for the outcomes.⁸¹⁵ I have argued that the contrary is true. Indeed, the ascription of legal personhood to certain autonomous AI systems—and I have focused on LAWs—would enable the ascription of liability vicariously to the organisations which deploy them, ensuring the appropriate legal expression and enforceability of the reparatory moral duties which accrue to those organisations. Secondly, one might ask: why stop at LAWs? After all, autonomous AI is predicted to take on an increasing number and variety of ‘jobs’ within a range of enterprises. Why not extend the legal personification of autonomous AI to ensure the appropriate enforceability of reparatory moral duties in all cases where the enterprise’s activities in the use of autonomous AI generate characteristic risks of harm to others? I am sympathetic to this line of thinking and suggest that vicarious liability is likely to be an appropriate mechanism for ensuring that those enterprises which use autonomous AI bear appropriate liability when the characteristic risks of that technology materialise. But further elaboration of this idea would be outside the scope of this dissertation.

9.3. Product liability for LAWs

In 9.1, I argued that it would satisfy the legislative principle of integrity to hold military organisations vicariously liable for the wrongful ‘conduct’ of LAWs in cases like Massacre,

⁸¹⁵ For examination and critique of this view: Turner (2019: 3.6.2).

provided that LAWs are capable of bearing primary liability for that ‘conduct’. In 9.2, I argued that we have independent reasons for ascribing legal personhood and liability to LAWs in cases like Massacre as part of giving legal expression and enforceability to the moral rights of civilians. So, I submit that the legal personification of LAWs is part of a promising solution to the ‘gap’ in private law mechanisms for the ascription of liability in cases like Massacre. This section sketches the rest of the solution. In essence, I submit that it would also satisfy the legislative principle of integrity to hold the manufacturers and sellers of LAWs liable under a modified doctrine of strict product liability for the outcome in cases like Massacre.⁸¹⁶ 9.3.1 addresses the criterion of soundness: in brief, I show that a core justificatory rationale for the development of the doctrine of strict product liability is to ensure the legal expression and enforceability of reparatory moral duties associated with enterprise liability, particularly in profit-making enterprises. 9.3.2 addresses the criterion of fit. I argue for targeted amendments to the doctrine of strict product liability to ensure the legal expression and enforceability of the moral rights of victims in cases like Massacre. These proposals fit with ongoing legal reforms in the EU and the US.

9.3.1. Product liability and enterprise liability

It is generally accepted that enterprise liability is at least part of the justificatory rationale for imposing strict product liability.⁸¹⁷ In the US, *Henningsen v. Bloomfield Motors*⁸¹⁸ is a landmark case in ensuring that manufacturers bear responsibility for injuries from a defective

⁸¹⁶ Liability could be appropriately distributed among the various enterprises which participate in the production and sale of LAWs through a mechanism like that of ‘contribution’ in English and Welsh law whereby a defendant to a private law action can bring a claim in contribution against a third party which shares liability for the cause of that action.

⁸¹⁷ The Law Commission and the Scottish Law Commission (1977) (English, Welsh and Scots law); Priest (1985); Keating (2017); Brodie (2010: 3-7) (US law); Stapleton (1986) (English, Welsh and US law); Whittaker (2005) (English, Welsh and French law).

⁸¹⁸ [1960] 32 N.J. 358.

product, regardless of negligence or a direct contract with the injured person. The *ratio* invokes the fairness idea underpinning enterprise liability:

‘the burden of losses consequent upon use of defective articles [should be] borne by those who are in a position to either control the danger or make an equitable distribution of the losses when they do occur.’⁸¹⁹

In *Greenman v. Yuba Power Products, Inc.*,⁸²⁰ Traynor J elaborated on the rationale for strict product liability:⁸²¹

‘to insure [*sic*] that the costs of injuries resulting from defective products are borne by the manufacturers that put such products on the market rather than by the injured persons who are powerless to protect themselves.’⁸²²

Indeed, in *Sindell v Abbott Laboratories*,⁸²³ the court approved of Sheiner’s theory⁸²⁴ that enterprise liability may support joint and several strict product liability for all members of a chain of distribution:

where ‘an entire industry, engaged in a predictably dangerous enterprise and following similar safety practices, places an identically defective product in the stream of commerce, the industry rather than the individual manufacturer should be the focal point for liability because it can best allocate risks, distribute costs and take preventive measures.’⁸²⁵

⁸¹⁹ *Ibid* at 379.

⁸²⁰ [1963] 59 Cal.2d 57.

⁸²¹ *C.f. Escola v. Coca Cola* [1944] 150 P.2d 436.

⁸²² *Greenman v. Yuba Power Products* at 64.

⁸²³ 26 Cal. 3d 588.

⁸²⁴ Sheiner (1977).

⁸²⁵ *Ibid*: 1002.

Although enterprise liability has not always been a popular principle,⁸²⁶ its role as one of the key justificatory principles underlying US product liability law is uncontroversial.⁸²⁷

Following the US trend, in English, Welsh and Scots law, the Law Commission and the Scottish Law Commission in 1965 argued for the ‘imposition of strict liability for injuries resulting from defects in products that are put into circulation in the course of a business, and that the liability should rest primarily on the producer’.⁸²⁸ Stapleton identified this kind of idea as ‘enterprise liability’—broadly, ‘sheeting home losses to the enterprise which was responsible for them’⁸²⁹—and argued for its place as one of two theoretical bases for the development of ‘strict’ rather than ‘fault-based’ product liability in Europe and the US.⁸³⁰ Indeed, part of Stapleton’s critique of the 1985 Council of the European Communities *Council Directive on Liability for Defective Products* is that it inadequately implements that enterprise liability rationale.⁸³¹ The second theoretical base for the development of strict product liability which Stapleton identifies is ‘deterrence theory’, that is, the ‘economic theory’ according to which the identification of defendants and the extent of their liability is determined by the optimisation of a function for the allocation of resources at the least cost of accidents.⁸³² We have seen this overlapping rationale in the earlier cited passages from the US judgments but I will not examine it further here: given the stated idealising assumptions in *Massacre*, the use of LAWs does not appear to be a state of affairs which calls for deterrence or, indeed, a more efficient allocation of resources in terms of the costs of accidents. But I suggest that if legal scholarship is correct that at least part of the rationale for the development of strict product

⁸²⁶ For a classic critique: Priest (1985); for a response: Keating (2017).

⁸²⁷ For further discussion of the US context: Brodie (2010: ch.1).

⁸²⁸ The Law Commission and the Scottish Law Commission (1977: 11, [38]).

⁸²⁹ Stapleton (1986: 397).

⁸³⁰ *Ibid.* Note that the sense of ‘responsible’ which she invokes is more akin to answerability than blameworthiness: ‘[responsibility] does not depend on whether the risk of injury could have been foreseen, discovered, prevented or controlled, but only on the fact that it was a risk created by the defendant’ (*ibid.*).

⁸³¹ *Ibid.* As we will see *infra*, the new Directive (2024/2853) seems to better express the enterprise liability rationale.

⁸³² *Ibid.* 394, 396.

liability law over the last forty years is enterprise liability, then the application of strict product liability in cases like Massacre would satisfy the Dworkinian criterion of soundness.

9.3.2. Four reforms...

In the last chapter, we saw that there appear to be four principal concerns which have been raised for the application of strict product liability in cases like Massacre. Firstly, there has been uncertainty internationally concerning whether autonomous AI is a product or a service. Secondly, if it is a product, there is no clear consensus that the system would be ‘defective’ in cases like Massacre. Thirdly, given the dynamic, evolving nature of machine learning software, there is a question of whether a claimant in a case like Massacre would be able to establish on the balance of probabilities that any defect in the LAW was present at the time that the system left the relevant producer’s or seller’s hands and reached the claimant without substantial change. Fourthly, even if a claimant can show that there is a product with a defect which existed at the relevant time (e.g., that the software is a product and contained a particular bug at the point of sale with inadequate safeguards for its management), it may be hard for claimants to establish on the balance of probabilities that the defect was causally operative in the production of the damage in question given the issues of technical and procedural opacity.

This section adumbrates four proposals for the reform of strict product liability mechanisms at the international level which, if embraced by states through a declaration or a treaty for the regulation of AI systems, would help to ensure the legal expression and enforceability of those reparatory moral duties which accrue to participants in enterprises for the production and sale of LAWs. In each case, I explain how these reforms at the international level would fit with the ongoing reforms to strict product liability mechanisms which are emerging locally in the US and European contexts.

9.3.2.1. Product

It should be stipulated that LAWs are ‘products’ and that all those enterprises which participate⁸³³ in their production and sale are ‘manufacturers’ for the purposes of product liability law.

This would fit with ongoing reforms in the EU context. Indeed, the new EU Product Liability Directive of 23 October 2024, which requires member states to ‘bring into force the laws, regulations and administrative provisions necessary to comply [...] by 9 December 2026’,⁸³⁴ expressly provides for a broadened definition of a ‘product’, to accommodate AI systems, and a broadened definition of a ‘manufacturer’, to encompass the increased complexity of supply chains.⁸³⁵ Under the new Directive, a product ‘can be tangible or intangible’, including ‘[s]oftware, such as operating systems, firmware, computer programs, applications or AI systems [...] irrespective of the mode of its supply or usage, and therefore irrespective of whether the software is stored on a device, accessed through a communication network or cloud technologies, or supplied through a software-as-a-service model’.⁸³⁶ Additionally, a ‘manufacturer’ includes any ‘developer or producer of software, including AI system providers’.⁸³⁷ This ‘significantly increases the liability exposure not only of manufacturers and importers, but also of additional actors involved in the selling or distribution of products.’⁸³⁸ The aim is to ensure ‘legal certainty’ in the application of ‘no-fault liability’ to software developers and producers.⁸³⁹

The approach which I propose at the international level would also fit with recent legal developments in the US. A key case is *Garcia v. Character.ai*, one of multiple lawsuits against

⁸³³ See *Garcia v. Character.ai* below for the specification that ‘participation’ should be ‘substantial’. I take no stand on the precise degree of participation which would be required to justify ‘manufacturer’-status.

⁸³⁴ EU Product Liability Directive (2024/2853), Art.22.

⁸³⁵ *Preamble*, EU Product Liability Directive (2024/2853) at (13).

⁸³⁶ *Ibid.* Information, the Directive notes, is not, however, a product.

⁸³⁷ *Ibid.*

⁸³⁸ Taylor Wessing (2025).

⁸³⁹ EU Product Liability Directive (2024/2853) at (13).

Google and the start-up Character.ai by the families of teenagers who died by suicide or self-harmed after interacting with chatbots on the platform Character A.I.⁸⁴⁰ In her ruling on 20 May 2025 in *Garcia v. Character.ai*, Judge Conway examined the ‘product’/‘service’ distinction under Florida law and held that:

‘Character A.I. is a product for the purposes of Plaintiff’s product liability claims so far as Plaintiff’s claims arise from defects in the Character A.I. app rather than ideas or expressions within the app.’⁸⁴¹

She also upheld the allegations against Google as a ‘component manufacturer’ of the platform notwithstanding that the platform was developed and operated by another company at the time of the teenager’s death.⁸⁴² This is because there was a plausible case that Google had ‘substantially participated’ in the integration of its LLMs into Character A.I.⁸⁴³ The proposal which I have sketched also fits with the US Aligning Incentives for Leadership, Excellence, and Advancement in Development (AI LEAD) Act of 29th September 2025, which classifies AI systems as products for the purposes of strict product liability in the US and expressly extends strict product liability for ‘unreasonably dangerous or defective’ AI systems⁸⁴⁴ to any ‘person who designs, codes, produces, owns, or substantially modifies’ an AI system.⁸⁴⁵ The first reform of international law which I have proposed therefore fits with the expansion of strict product liability mechanisms to accommodate AI systems and their producers and sellers in the EU and the US.

9.3.2.2. Defect

⁸⁴⁰ The parties to *Garcia v. Character.ai* and the other lawsuits have agreed to settle: Criddle (2026).

⁸⁴¹ *Garcia v. Character.ai* at 36.

⁸⁴² *Ibid* at 18-24.

⁸⁴³ *Ibid* at 19.

⁸⁴⁴ AI LEAD Act at §101.

⁸⁴⁵ *Ibid* at §3.

Secondly, there needs to be a rebuttable presumption under international private law that cases like Massacre involve a defective product. One way to think about this is in light of the idea of ‘obvious malfunction’ under the new EU Directive.⁸⁴⁶ For example, where ‘a glass bottle [...] explodes in the course of reasonably foreseeable use, [...] it is unnecessarily burdensome to require a claimant to prove defectiveness when the circumstances are such that its existence is undisputed.’⁸⁴⁷ This echoes a classic US product liability case, *Escola v. Coca Cola*, where the court seems to have treated exploding Coca Cola bottles as presumptively defective. I suggest that a similarly common-sense approach should apply in cases like Massacre: the LAW fails to function as intended during reasonably foreseeable use, causing serious damage; the presumption should be that the product was defective. One might object that warnings could state that there is a risk of events like Massacre, thereby shifting responsibility onto the military which procures the LAW.⁸⁴⁸ But this occludes the principle of enterprise liability. Consider:

Fancy gun: P buys a large number of high-precision guns for P’s military operations. Their use promises to enhance P’s military advantage over a region. P supplies P’s soldiers with the guns and they line up to practice. One of the guns explodes, killing the soldier holding it.

Fancy gun (warning): P buys a large number of high-precision guns for P’s military operations. Their use promises to enhance P’s military advantage over a region. The manufacturer warns that there is a very small possibility of guns exploding. P supplies P’s soldiers with the guns and they line up to practice. One of the guns explodes, killing the soldier holding it.

⁸⁴⁶ EU Product Liability Directive (2024/2853) at (46).

⁸⁴⁷ *Ibid.*

⁸⁴⁸ C.f. Corn (2014).

I suggest that the manufacturer's warning in Fancy gun (warning) should not be enough to remove liability for the obvious defect in the product. There is an enterprise which has created a characteristic risk of a bad outcome. That enterprise should internalise the cost of that characteristic risk. A warning is necessary but not sufficient to fairly distribute risk.

9.3.2.3. Defective at the point of sale

It follows from this identification of AI systems as products for the purposes of strict product liability that the capacity of the product to change through 'learning' after leaving the hands of a manufacturer is a risk characteristic of that particular kind of product. Enterprises which produce and sell these products should internalise that risk, first, through introducing adequate safeguards over the learning software. If a machine learning product is found to be defective, I suggest that a rebuttable presumption should apply that the defect was present at the point of sale and reached the claimant without substantial change. This reverses the burden of proof as to the second limb of the standard test for strict product liability. The motivation for this reversal is to incentivise transparency and disclosure. Indeed, a claimant would otherwise be required to show in practice that a defect was not learnt after the point of sale, which would create a perverse incentive on manufacturers to market (technically and procedurally) opaque products and to minimise disclosure so as to evade liability for poor safeguards. This presumption of a defect at the point of sale would be rebutted by a manufacturer's evidencing that a defect was learnt after sale due to unreasonable use or that the buyer substantially changed the product.

This reversal of the burden of proof as to the second limb of the strict product liability test would fit with recent reforms in the US and the EU. In the US, the AI LEAD Act aims to 'ensure[...] that AI companies are incentivized to design their systems with safety as a priority' and 'to consider safety throughout the design and development process of AI products; not only

when products fail and things go wrong’.⁸⁴⁹ Besides safeguards throughout the life cycle of a learning product, companies should ‘[i]mplement[...] transparent documentation of AI decision-making’.⁸⁵⁰ Similarly in the EU, onerous disclosure duties are imposed on manufacturers under Article 9 of the new Directive; Article 10(2)(a) provides that failure to comply with the disclosure duty triggers a rebuttable presumption that the product is defective. This presumption is rebuttable in circumstances like those glossed above.⁸⁵¹ The idea in the EU and the US is to make opacity expensive, whether it is technical or procedural. And this too I suggest reflects enterprise liability: the extent of opacity is for an enterprise to choose but it should be at least as expensive as compensating the victims of the lethal accident. In this way, the cost of non-disclosure reflects the cost of compensating the victims. This protects claimants from becoming victims of the compounded injustices of non-compensation through non-disclosure of information deemed sensitive. Indeed, the enterprise which withholds information for commercial gain should internalise the costs of that business activity, costs which include compensating the victims of lethal accidents involving LAWs.

9.3.2.4.Causation

Finally, for similar reasons as above, there should be a rebuttable presumption in obvious malfunction cases of a causal link between the defectiveness of the product and the damage. This mitigates the intractable difficulties for claimants in seeking to establish the causal history of a lethal accident where part of that history involves the opaque decision-making system of an autonomous AI.

Indeed, this presumption would fit with similar moves in the US and the EU. In the US AI LEAD Act, §101 provides that a causal link between an AI system defect and a resulting

⁸⁴⁹ US Senate Committee on the Judiciary (2025).

⁸⁵⁰ Young (2025).

⁸⁵¹ EU Product Liability Directive (2024/2853) at (39).

injury may be inferred without proof of a specific defect ‘when the incident that harmed the claimant—(A) was of a kind that ordinarily occurs as a result of [an AI system] defect; and (B) was not, in the particular case, solely the result of [other] causes.’⁸⁵²

In the EU, the new Directive explicitly recognises that claimants are ‘often at a significant disadvantage compared to manufacturers in terms of access to, and understanding of, information on how a product was produced and how it operates’⁸⁵³ such that ‘[w]here it has been established that a product is defective and the kind of damage that occurred is, based primarily on similar cases, typically caused by the defectiveness in question, the claimant should not be required to prove the causal link and its existence should be presumed’.⁸⁵⁴

In the UK context, the Law Commission has recently set out to assess whether the Consumer Protection Act, which implemented the earlier European Product Liability Directive, should be updated along similar lines to the new Directive. In this evolving legal context, I suggest that the four modest reforms which I have sketched could form part of an international agreement for the regulation of LAWs and would fit with ongoing developments in strict product liability law in the US and Europe.

⁸⁵² AI LEAD Act at 10-11.

⁸⁵³ *Ibid* at (42).

⁸⁵⁴ *Ibid* at (47).

10. Closing remarks on the Responsibility Gaps Thesis

If we don't take action now

*We settle for nothing later*⁸⁵⁵

In Part Two of this dissertation, I have sought to make three principal contributions to the AI ethics literature which I will briefly recapitulate. Firstly, I argued that a thesis concerning responsibility for AI systems which has dominated the literature on LAWs for the last twenty years—viz., the Responsibility Gaps Thesis—is false, at least as it is usually understood. This is important because LAWs are coming and it is already too easy, as per Dylan’s ‘Masters of War’, for those responsible for killing in war to ‘hide behind walls [and] hide behind desks.’⁸⁵⁶

I distinguished two popular variants of the Responsibility Gaps Thesis. The first rests on the presumption that it is better, other things being equal, if someone is responsible for a bad outcome than if no one is. I showed in chapter six that that presumption is false. Chapter seven considered an alternative interpretation of the Responsibility Gaps Thesis according to which cases of purported ‘responsibility gaps’, like Massacre, are alleged to be counterexamples to our best theories of moral responsibility. The idea is that they demonstrate a discrepancy between how much moral responsibility is attributable according to our best theories and how much moral responsibility *should* be attributable according to our intuitions. I showed that this interpretation of the thesis is false: indeed, it seems to trade on an equivocation between two distinct senses of ‘moral responsibility’, namely answerability and blameworthiness. According to the theory of moral responsibility as answerability which I defended, lots of agents are answerable for Massacre; according to moral responsibility as

⁸⁵⁵ Rage Against The Machine (1992).

⁸⁵⁶ Dylan (1963).

blameworthiness, no one is blameworthy for Massacre. Both results turned out to be intuitively right given the idealising assumptions. So there are no ‘gaps’ in our best theories of moral responsibility.

Secondly, I articulated and defended a third, alternative interpretation of the Responsibility Gaps Thesis according to which there is some discrepancy between the (degree of) legal responsibility and the (degree of) moral responsibility which is attributable in cases like Massacre. I argued that the Responsibility Gaps Thesis so interpreted is false as a statement about ‘gaps’ in criminal law since punishment would serve no intrinsic or instrumental good in alleged ‘responsibility gaps’ cases and, quite rightly, no one bears liability under existing criminal law mechanisms. However, I argued that the Responsibility Gaps Thesis is true as a statement about the moral adequacy of international private law remedies in cases like Massacre. I contended that those who participate in enterprises which create characteristic risks to others bear reparatory moral duties when those risks materialise. The production, sale and use of LAWs create characteristic risks to civilians, namely risks of cases like Massacre. As such, participants in the enterprises which produced, sold and deployed the LAW in cases like Massacre bear reparatory moral duties to the victims and their families. However, those duties do not seem to be reflected in existing private law mechanisms. So, I highlighted a novel ‘responsibility gap’: how, in principle, can we give legal expression and enforceability in private law to those reparatory moral duties which accrue to participants in enterprises for the production, sale and use of the LAW in cases like Massacre?

Thirdly, in chapter nine, I offered a novel solution to that ‘responsibility gap’ which I identified in chapter eight. Firstly, I argued for the application of vicarious liability in cases like Massacre in order to give legal expression and enforceability to the reparatory moral duties which accrue to participants in the military enterprise which deploys the LAW. A corollary of this is that LAWs should be legally personified. I articulated and defended a novel argument

for the legal personhood of LAWs and the ascription to LAWs of a legal duty not to target civilians. Finally, I argued for four targeted reforms to strict product liability mechanisms to ensure the legal expression and enforceability of the reparatory moral duties which accrue in cases like Massacre to participants in the enterprises which produced and sold the LAW. That completes my response to the Responsibility Gaps Thesis.

CONCLUSION

*Arise! For this matter is your responsibility, but we will be with you; be courageous and act.*⁸⁵⁷

This dissertation has propounded two novel arguments for the legal personification of certain autonomous AI systems. Part One argued that the legal personification of SARs would form part of an appropriate legal response to the risk of robotic sexual violence which the imminent proliferation of SARs introduces. The ascription of legal rights to SARs against subjection to robotic sexual violence is justified by appeal to the independent value of reciprocity since robotic sexual violence undermines measures designed to ensure adequate all-purpose means for all citizens to make effective use of their basic freedoms. Part Two argued that the legal personification of LAWs would form part of an appropriate legal response to the risk of lethal accidents which the imminent proliferation of LAWs introduces. The ascription of a legal duty to LAWs not to direct attacks against civilians is justified by appeal to the role which LAWs play as quasi-employees of a military enterprise. In addition, I argued for four targeted reforms to strict product liability mechanisms, which reflect the reparatory moral duties which accrue in lethal accidents cases to participants in the enterprises which produced and sold the LAW.

A closing note on methodology: this dissertation adopted a challenge-based methodology, that is, one which focuses on specific challenges in the literature and develops novel solutions as opposed to articulating general principles for application to specific cases. The key virtue as well as the key shortcoming of this methodology, to which the Introduction adverted, is that the conclusions are particular to a given challenge. My dissertation therefore

⁸⁵⁷ Ezra (10:4).

addresses legal personhood in two particular contexts—that of robotic sexual violence involving SARs and that of lethal accidents involving LAWs—and purports to make no definitive claims about legal personhood for autonomous AI in general. That said, the arguments advanced in this dissertation shine light on an original and suggestive general idea, a proof of principle, that the ascription of legal personhood to certain kinds of autonomous AI systems is morally justified notwithstanding that these systems have no intrinsic moral status. Indeed, such ascriptions may be necessary for the advancement of justice and fairness. This is not an idea which, to my knowledge, has been developed by any other author working on AI ethics in the law. It is therefore my primary contribution to that literature.

References

Bibliography

Abbas, A., 2025, 'The AI Feedback Loop: When Machines Amplify Their Own Mistakes by Trusting Each Other's Lies', *Unite AI* (15th May 2025). Available online: <https://www.unite.ai/the-ai-feedback-loop-when-machines-amplify-their-own-mistakes-by-trusting-each-others-lies/> (accessed: 4th January 2026).

Abbey, R., 2011, *The Return of Feminist Liberalism*, Taylor & Francis Group

Abbott, R., 2018, 'The Reasonable Computer: disrupting the Paradigm of Tort Liability', *The George Washington Law Review*, Vol. 86(1).

Abbott, R., and Sarch, A., 2019, 'Punishing Artificial Intelligence: Legal Fiction and Science Fiction', *University of California, Davis* 53: 323

Ackerman, B., 1980, *Social Justice and the Liberal State*, New Haven: Yale University Press

Adamu, A., 2020, 'Analyses of the challenges faced by the international criminal court in the exercise of its jurisdiction', *International Journal of Law* Vol.6(6):98-105. Available online: <https://www.lawjournals.org/assets/archives/2020/vol6issue6/6-5-79-412.pdf> (accessed: 20th January 2026).

Adarsh, H. and Sahoo, S., 2023, 'Pornography and Its Impact on Adolescent/Teenage Sexuality', *Journal of Psychosexual Health* 5(1)

Additional Protocol I, 1977, Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of International Armed Conflicts, June 8 1977 1125 U.N.T.S. 3

Adler, P., Falk, C., Friedler, S. A., Rybeck, G., Scheidegger, C., Smith, B., Venkatasubramanian, S., 2016, 'Auditing Black-box Models for Indirect Influence', IEEE International Conference

on Data Mining (ICDM), 2016. Available online: <https://arxiv.org/abs/1602.07043> (accessed: 19th January 2026)

Alberts, P., Calfo, P. & Gabat, J., 2025, 'Artificial Intelligence: The 'Black Box' of Product Liability', Husch Blackwell (Apr. 4, 2025). Available online: <https://www.productlawperspective.com/2025/04/artificial-intelligence-the-black-box-of-product-liability/> (accessed: 20th January 2026).

Aligning Incentives for Leadership, Excellence, and Advancement in Development (AI LEAD) Act 2025. Available online: <https://www.govtrack.us/congress/bills/119/s2937/text/is> (accessed: 20th January 2026).

Altwickler, T., 2015, 'Rechtsperson im Rechtspositivismus' in R. Gröschner, S. Kirste, O. Lembcke (eds.), *Person und Rechtsperson: zur Ideengeschichte der Personalität*, Mohr Siebeck

Amoroso, D., 2020, *Autonomous Weapons Systems and International Law: A Study on Human-Machine Interactions in Ethically and Legally Sensitive Domains* (Edizioni Scientifiche Italiane, 2020).

Amoroso, D. and Tamburrini, G., 2021a, 'In Search of the 'Human Element': International Debates on Regulating Autonomous Weapons Systems', *The International Spectator*, 56(1):20–38. Available online: <https://doi.org/10.1080/03932729.2020.1864995> (accessed: 21st January 2021).

—, 2021b, 'Toward a Normative Model of Meaningful Human Control over Weapons Systems', *Ethics & International Affairs* 35, no. 2: 245-272.

Anderson, E., 1991, 'John Stuart Mill and Experiments in Living', *Ethics* 102: 4–26

Anderson, M., and Anderson, S., 2011, 'General Introduction,' in M. Anderson and S. Anderson (eds.) *Machine Ethics*, Cambridge: Cambridge University Press: 1–4

Anthropic, 2023, 'Claude's Constitution', May 9 2023, <https://www.anthropic.com/index/claudes-constitution> (accessed: 19th January 2026)

Aristotle, 2009, *The Nicomachean Ethics*, Ross, D. (trans.), Oxford University Press: Oxford World's Classics

Arkin, R., 2009, *Governing Lethal Behavior in Autonomous Robots*, Boca Raton: CRC Press

—, 2014, 'The Case for Ethical Autonomy in Unmanned Systems' in T. Demy *et al.* (eds.) *Military Ethics and Emerging Technologies*, Routledge.

Arkoudas, K., Bringsjord, S. and Bello, P., 2005, 'Toward ethical robots via mechanized deontic logic', in *Machine ethics: Papers from the 2005 AAAI fall symposium*: 17-23

Article 36, 2016, 'Key elements of meaningful human control', Geneva 11-15 April 2016. Available online: <https://www.article36.org/wp-content/uploads/2016/04/MHC-2016-FINAL.pdf> (accessed: 19th January 2026).

Asaro, P., 2012, On Banning Autonomous Weapon Systems: Human Rights, Automation, and the Dehumanization of Lethal Decision-Making, 94 INT'L REV. RED CROSS 687, 701 (2012);

—, 2016, 'Jus nascendi, Robotic Weapons and the Martens Clause' in R. Calo, M. Froomkin and I. Kerr (eds.) *Robot Law*, Cheltenham UK: Edward Elgar Publishing: 367-386

—, 2020, 'Autonomous Weapons and the Ethics of Artificial Intelligence', in S. M. Liao (ed.) *Ethics of Artificial Intelligence*, Oxford University Press 2020

Asimov, I., 1950, 'Runaround', reprinted in *I, Robot*, New York: Doubleday

Atiyah, P., 1967, 'Vicarious Liability in the Law of Torts', Butterworths.

Audi, R., 2006, 'Testimony, credulity, and veracity', *The Epistemology of Testimony*, J. Lackey and E. Sosa (eds.), Oxford: Oxford University Press: 25–49

Austrian Civil Code, Available online: https://www.trans-lex.org/602100/_/austrian-civil-code/ (accessed: 19th January 2026).

Autonomous Weapons, 2025a, 'Milestones in the Global Legal Framework for Autonomous Weapons', *Autonomous Weapons*, 22 April 2025. Available online: <https://autonomousweapons.org/global-legal-framework-milestones/> (accessed: 19th January 2026).

—, 2025b, 'The Political Landscape: How Nations are Responding to Autonomous Weapons in War', *Autonomous Weapons*, 25 April 2025. Available online: <https://autonomousweapons.org/global-perspectives-on-regulation/> (accessed: 19th January 2026).

BAE 'Taranis'. Available online: <https://www.baesystems.com/en/product/taranis> (accessed: 5th January 2026).

Baehr, A., 2008, 'Perfectionism, Feminism and Public Reason', *Law and Philosophy* 27: 193–222

Bagshaw, R., 2024, 'What is 'Reasonable Foreseeability'?', in K. Burns, and others (eds), *Torts on Three Continents: Honouring Jane Stapleton* (Oxford, 2024; online edn, Oxford Academic, 19 Sept. 2024), <https://doi-org.ezproxy-prd.bodleian.ox.ac.uk/10.1093/oso/9780198889748.003.0008> (accessed: 8th January 2026).

Barry, C. and Wiens, D., 2016, 'Benefiting from wrongdoing and sustaining wrongful harm', *Journal of Moral Philosophy* 13.

Bartneck, C., van der Hoek, M., Mubin, O., and Al Mahmud, A., 2007, '*Daisy, Daisy, Give me your answer do!*' – *Switching off a robot*, PROCEEDINGS OF THE 2ND ACM/IEEE INTERNATIONAL CONFERENCE ON HUMAN-ROBOT INTERACTION, Washington DC: 217–22

Bathae Y., 2018, 'The artificial intelligence black box and the failure of intent and causation', *Harv J Law Technol* 31(2):889–938

Bayern, S., 2016, 'The Implications of Modern Business-Entity Law for the Regulation of Autonomous Systems', *European Journal of Risk Regulation* 2: 297-309

BBC, 2020, 'Uber's self-driving operator charged over fatal crash'. Available online: <https://www.bbc.co.uk/news/technology-54175359> (accessed: 19th January 2026).

BBC, 2021, 'Tesla Autopilot: US opens official investigation into self-driving tech'. Available online: <https://www.bbc.co.uk/news/technology-58232137> (accessed: 19th January 2026).

Behrendt, P. and Chandler, K., 2025, 'AI liability – who is accountable when artificial intelligence malfunctions?', *Taylor Wessing* 7 January 2025. Available online: <https://www.taylorwessing.com/en/insights-and-events/insights/2025/01/ai-liability-who-is-accountable-when-artificial-intelligence-malfunctions> (accessed: 20th January 2026).

Bellaby, R. W., 2021, 'Can AI Weapons Make Ethical Decisions?', *Criminal Justice Ethics* 40(2): 86-107

Bellamy, A. J., 2008, 'The Responsibilities of Victory: 'Jus Post Bellum' and the Just War', *Review of International Studies*, 34(4): 601–625. Available online: <http://www.jstor.org/stable/40212494> (accessed: 19th January 2026).

Bellon, T., 2018, 'Fatal U.S. self-driving auto accident raises novel legal questions' *Reuters*. Available online: <https://www.reuters.com/article/technology/fatal-us-self-driving-auto-accident-raises-novel-legal-questions-idUSKBN1GW2VX/>. (accessed: 19th January 2026).

Benhabib, S., 2002, *The Claims of Culture: Equality and Diversity in the Global Era*, Princeton: Princeton University Press

Berg, J., 2007, 'Of Elephants and Embryos: A Proposed Framework for Legal Personhood', 59 *Hastings Law Journal*

Berk, R. A. and Bleich, J., 2013, 'Statistical Procedures for Forecasting Criminal Behaviour: A Comparative Assessment', *Criminology & Public Policy* 12(3)

Berman, M. N., 2024, 'Retributivism', in J. Ryberg (ed.), *The Oxford Handbook of the Philosophy of Punishment*, Oxford Handbooks (2024; online edn, Oxford Academic, 22 Oct. 2024). Available online: <https://doi.org/10.1093/oxfordhb/9780197750506.013.3> (accessed: 20th January 2026).

Bianchi, A., 2020, 'Considering sex robots for older adults with cognitive impairments', *Journal of Medical Ethics* Vol. 47, No. 1

Bilchitz, D., 2009, 'Moving Beyond Arbitrariness: The Legal Personhood and Dignity of Non-Human Animals', *South African Journal on Human Rights* 25(1): 38-72

Billingham, P. and Taylor, A., 2023, 'Can Civic Friendship Ground Public Reason?', *The Philosophical Quarterly* Vol 0, No. 0: 1-22

Bird, C., 2014, 'Coercion and Public Justification,' *Politics, Philosophy, & Economics*, 13(3): 189–214

Blackstone, W., 1876, 'The commentaries on the laws of England of Sir William Blackstone', London : John Murray, Available online:

<https://archive.org/details/commentariesonla01blacuoft/page/viii/mode/2up> (accessed: 19th January 2026).

Bo, M., Bruun, L. and Boulanin, V., 2022, 'Retaining Human Responsibility in the Development and Use of Autonomous Weapon Systems: On Accountability for Violations of International Humanitarian Law Involving AWS', *SIPRI*, October 2022. Available online: https://www.sipri.org/sites/default/files/2022-10/2210_aws_human_responsibility.pdf (accessed: 19th January 2026).

Boden, M.A., 2016, *AI: Its Nature and Future*, Oxford University Press, New York.

Bohman, J., 1996, *Public Deliberation: Pluralism, Complexity, and Democracy*, Cambridge, MA: MIT Press

—, 2003, 'Deliberative Toleration,' *Political Theory*, 31(3): 757–779

Bonafé, B., 2009, 'Command Responsibility', in *THE OXFORD COMPANION TO INTERNATIONAL CRIMINAL JUSTICE*

Bostrom, N., 2012, 'The Superintelligent Will: Motivation and Instrumental Rationality in Advanced Artificial Agents', *Minds and Machines* 22 no. 2: 71-85

—, 2014, *Superintelligence: Paths, Dangers, Strategies*, Oxford: Oxford University Press

Bovens, M., 1998, *The quest for responsibility: Accountability and citizenship in complex organisations*, Cambridge University Press

Braham, M. and van Hees, M., 2012, 'An Anatomy of Moral Responsibility', *Mind* 121(483): 601-634

- Braithwaite, S. R., Coulson, G., Keddington, K. and Fincham, F. D., 2015, 'The Influence of Pornography on Sexual Scripts and Hooking Up Among Emerging Adults in College', *Archives of Sexual Behavior*, 44: 111–23
- Brennan, S., 2004, 'Liberalism, Feminism, and Rights', in *Varieties of Feminist Liberalism*, A. Baehr (ed.), Lanham MD: Rowman and Littlefield
- Brennan, T. and Oliver, W. L., 2013, 'The Emergence of Machine Learning Techniques in Criminology', *Criminology & Public Policy*, 12(3)
- Brennan, J. and Jaworski, M., 2015, 'Markets Without Symbolic Limits', *Ethics* 125(4): 1053-1077
- Brink, D., 2008, 'Mill's Liberal Principles and Freedom of Expression' in *Mill's 'On Liberty': A Critical Guide*, C. L. Ten (ed.), New York: Cambridge University Press
- Brodie, D., 2007, 'Enterprise Liability: Justifying Vicarious Liability', *Oxford Journal of Legal Studies*, 27(3):493–508. Available online: <http://www.jstor.org/stable/4494596> (accessed: 20th January 2026).
- , 2010, *Enterprise Liability and the Common Law*, Cambridge University Press.
- Brown, C., 2019, 'Sex Robots, Representation, and the Female Experience', *The American Papers* 37: 105–118
- Browne, J., 2023, 'AI and Structural Injustice: A Feminist Perspective', in J. Browne, and others (eds), *Feminist AI: Critical Perspectives on Algorithms, Data, and Intelligent Machines* (Oxford, 2023; online edn, Oxford Academic, 23 Nov. 2023). Available online: <https://doi.org/10.1093/oso/9780192889898.003.0019> (accessed: 20th January 2026).
- , 2025, *Political Responsibility and Tech Governance: AI, Repro-tech and Structural Injustice*, Cambridge University Press

- Bryan, G., 1999, *Black's Law Dictionary* (7th ed.). St. Paul, Minn: West Group
- , 2024, *Black's Law Dictionary* (12th ed.). Thomson Reuters.
- Bryson, J. J., Diamantis, M. E., & Grant, T. D., 2017, 'Of, for, and by the people: the legal lacuna of synthetic persons', *Artificial Intelligence and Law*, 25(3), 273–291.
- Buchan, R. and Tsagourias, N., 2020, 'Autonomous Cyber Weapons and Command Responsibility', 96 *International Law Studies* 645.
- Buchwald, E., Fletcher, P. R., Roth, M. (eds.), 2005, *Transforming a rape culture*, Milkweed editions, Minneapolis
- Buhbut, A., 2021, 'בלעדי | להקות הרחפנים של צה"ל השתתפו בלחימה בעזה, וחוקי המשחק צפויים להשתנות', Walla News June 5 2021, <https://news.walla.co.il/item/3439695> (accessed: 19th January 2026).
- Buiten, M., 2024, 'Product liability for defective AI', 57 *Eur. J. L. Econ.* 239. Available online: <https://link.springer.com/article/10.1007/s10657-024-09794-z> (accessed: 20th January 2026).
- Burge, T., 1993, 'Content preservation', *The Philosophical Review* 102: 457–488
- , 1997, 'Interlocution, perception, and memory', *Philosophical Studies* 86: 21–47
- Burge-Hendrix, B., 2007, 'The Educative Function of Law', *Law and Philosophy* M. Freeman and R. Harrison (eds.), OUP: 243-254
- Burri, S., 2017, 'What is the Moral Problem with Killer Robots?', in B. J. Strawser, R. Jenkins, and M. Robillard (eds.) *Who Should Die? The Ethics of Killing in War*, Oxford Scholarship Online
- Burrows, A. (ed.), 2017, *English Private Law*, (3rd edn. Oxford: Oxford University Press, 2017)

- Butt, D., 2007, 'On Benefiting from Injustice', *Canadian Journal of Philosophy*, 37(1):129–152. Available online: <http://www.jstor.org/stable/40232312> (accessed: 20th January 2026).
- , 2014, 'A Doctrine Quite New and Altogether Untenable', *J Appl Philos*, 31:336–348. Available online: <https://doi.org/10.1111/japp.12073> (accessed: 20th January 2026).
- , 2020 'The ethical implications of benefiting from injustice', in H. LaFollette (ed.), *Ethics in Practice: An Anthology (5th edition)*, Wiley Blackwell
- , 2021, 'What structural injustice theory leaves out', *Ethical Theory and Moral Practice* 24, 1161–1175. Available online: <https://rdcu.be/citcZ> (accessed: 20th January 2026).
- Campbell, J., 2020, *Causation in Psychology*, Harvard University Press
- Cane, P., 1993, *Atiyah's Accidents, Compensation and the Law*, London: Butterworths.
- Caney, S., 2006, 'Environmental Degradation, Reparations, and the Moral Significance of History', *Journal of social philosophy*, Vol.37 (3):464–482
- , 2010, 'Climate Change and the Duties of the Advantaged', *Critical Review of International Social and Political Philosophy* 13(1): 203–228
- Carline, A., 2011, 'Criminal justice, extreme pornography and prostitution: Protecting women or promoting morality?', *Sexualities* 14(3): 312–333
- Carpenter, J., 2013, *THE QUIET PROFESSIONAL: AN INVESTIGATION OF U.S. MILITARY EXPLOSIVE ORDNANCE DISPOSAL PERSONNEL INTERACTIONS WITH EVERYDAY FIELD ROBOTS*, Dissertation, University of Washington
- Cassese, A., 1998, 'Reflections on International Criminal Justice,' *Modern Law Review*, vol. 61.
- , 2013, *INTERNATIONAL CRIMINAL LAW* 65–66 (3d ed. 2013)

Catullus, 5, 'Let us live'. Available online: https://www.poetryintranslation.com/PITBR/Latin/Catullus.php#anchor_Toc531846733

(accessed: 19th January 2026).

CCW, Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons which may be Deemed to be Excessively Injurious or to Have Indiscriminate Effects.

CEDAW, 1979, 'Convention on the Elimination of All Forms of Discrimination against Women', Available online: <https://www.ohchr.org/en/instruments-mechanisms/instruments/convention-elimination-all-forms-discrimination-against-women>

(accessed: 19th January 2026).

Chambers, C., 2018, *Against Marriage: An Egalitarian Defense of the Marriage-Free State*, Oxford: Oxford University Press

Chapman, T., 1989, 'Freedom Now', *Crossroads*. Available online: <https://genius.com/Tracy-chapman-freedom-now-lyrics> (accessed: 21st January 2026).

Char, D. S., Shah, N. H., & Magnus, D., 2018, 'Implementing machine learning in healthcare—Addressing ethical challenges', *New England Journal of Medicine*, 378, 981–983

Charter of Fundamental Rights of the European Union. Available online: https://eur-lex.europa.eu/eli/treaty/char_2012/oj/eng (accessed: 20th January 2026).

Chatterjee, B. B., 2020, 'Child sex dolls and robots: challenging the boundaries of the child protection framework', *International Review of Law, Computers & Technology*, Vol. 34, No.1: 22-43

Chen, W., Yaguchi, Y., Naruse, K., Watanobe, Y., Nakamura, K., & Ogawa, J., 2018, 'A study of robotic cooperation in cloud robotics: Architecture and challenges', *IEEE Access*, 6, 36662–

36682. Available online: <https://doi.org/10.1109/ACCESS.2018.2852295> (accessed: 20th January 2026).

Chengeta, T., 2016, 'Accountability Gap: Autonomous Weapon Systems and Modes of Responsibility in International Law', *Denver Journal of International Law & Policy* Vol.45(1), Article 3. Available online: <https://www.scispace.com/pdf/accountability-gap-autonomous-weapon-systems-and-modes-of-28tgbof64l.pdf> (accessed: 19th January 2026).

Chesterman, S., 2020, 'Artificial Intelligence and the Limits of Legal Personality', *International & Comparative Law Quarterly* 69(4): 819-844

Chopra, S. & White, L. F., 2011, *A legal theory for autonomous artificial agents*, Ann Arbor: University of Michigan Press

Coffee, J. C., Jr., 1991, Does 'Unlawful' Mean 'Criminal'? Reflections on the Disappearing Tort/Crime Distinction in American Law, 71 B.U. L. REV. 193

Coeckelbergh M., 2020, 'Artificial intelligence, responsibility attribution, and a relational justification of explainability', *Sci Eng Ethics* 26(4):2051–2068. Available online: <https://doi.org/10.1007/s11948-019-00146-8> (accessed: 19th January 2026).

Cohen, J., m.s., 'Don't Shoot the Algorithm', personal communication

Collingridge, D., 1980, *The Social Control of Technology*, London: Frances Pinter.

Collins, S., 2019, 'Collective Responsibility Gaps', *Journal of Business Ethics*, 154(4):943–954.

Commonwealth of Australia, 2009, *Australian Government Attorney-General Department Discussion Paper: 'Should the Australian National Classification Scheme include an R18+ classification category for computer games?'*, Attorney General's Department, Canberra

Conard, A., 1976, *Corporations in Perspective*, Foundation Press

Consumer Protection Act 1987. Available online: <https://www.legislation.gov.uk/ukpga/1987/43/contents> (accessed: 20th January 2026).

Corn, G., 2014, Autonomous Weapon Systems: Managing the Inevitability of ‘Taking the Man Out of the Loop’ 21 (unpublished manuscript). Available online: <https://perma.cc/9KMQ-2VGQ> (accessed: 19th January 2026).

Cowen, N., 2015, ‘Millian Liberalism and Extreme Pornography’, *American Journal of Political Science* 60(2): 509-520

CPS (Crown Prosecution Service), 2021, ‘Tackling Rape Myths and Stereotypes’, in *Legal Guidance, Rape and Sexual Offences*: chapter 4. Available online: <https://www.cps.gov.uk/legal-guidance/rape-and-sexual-offences-chapter-4-tackling-rape-myths-and-stereotypes> (accessed: 19th January 2026).

Crépault, C. and Couture, M., 1980, ‘Men’s Erotic Fantasies’, *Archives of Sexual Behaviour* 9

Criddle, C., 2026, ‘Character.ai and Google agree to settle lawsuits over teen suicides’, *FT* 7 January 2026. Available online: <https://www.ft.com/content/ac518567-d901-4fae-86a3-eab54b12a81d> (accessed: 20th January 2026).

Criminal Justice and Immigration Act, 2008. Available online: <https://www.legislation.gov.uk/ukpga/2008/4/part/5/crossheading/pornography-etc> (accessed: 19th January 2026).

Critelli, J. and Bivona, J., 2008, ‘Women’s Erotic Rape Fantasies: An Evaluation of Theory and Research’, *Journal of Sex Research* 45

Crootof, R., 2015, ‘The Killer Robots Are Here: Legal and Policy Implications’, 36 *CARDOZO L. REV.* 1837

—, 2016, ‘War Torts: Accountability for Autonomous Weapons’, *University of Pennsylvania Law Review*, May 2016, Vol. 164, No. 6: 1347-1402

—, 2022, ‘War Torts’, *NEW YORK UNIVERSITY LAW REVIEW* 97(4): 1063-1142.

Curren, R. R., 1989, ‘The Contribution of *Nicomachean Ethics* III.5 to Aristotle’s Theory of Responsibility’, *History of Philosophy Quarterly* 6

Da Silva, M., 2024, ‘Responsibility gaps’, *Philosophy Compass*, e70002. Available online: <https://doi-org.ezproxy-prd.bodleian.ox.ac.uk/10.1111/phc3.70002> (accessed: 19th January 2026).

Dam, C. Van, 2013, ‘France’, *European Tort Law*, 2nd edn (Oxford, 2013; online edn, Oxford Academic, 23 May 2013). Available online: <https://doi.org/10.1093/acprof:oso/9780199672264.003.0003> (accessed: 20th January 2026).

Danaher, J., 2016a, ‘Robots, Law and the Retribution Gap’, *Ethics and Information Technology* 18(4): 299-309.

—, 2016b, ‘The threat of algocracy: Reality, resistance and accommodation’, *Philosophy and Technology*, 29(3): 245–268.

—, 2017a, ‘Robotic Rape and Robotic Child Sexual Abuse: Should They be Criminalised?’, *Criminal Law and Philosophy* 11: 71-95

—, 2017b, ‘The Symbolic Consequences Argument in the Sex Robots Debate’, in J. Danaher and N. McArthur (eds.) *Robot Sex: Social and Ethical Implications*, MIT Press, Cambridge

—, 2019, ‘Regulating Child Sex Robots: Restriction or Experimentation?’, *Med Law Rev.* 2019 Nov 1, 27(4): 553-575

——, 2022, ‘Tragic Choices and the Virtue of Techno-Responsibility Gaps’, *Philosophy & Technology* 35:26.

Darling, K., 2014[2016], ‘Extending legal protection to social robots: the effects of anthropomorphism, empathy, and violent behavior towards robotic objects’ in M. Froomkin, R. Calo, I. Kerr (eds.), *Robot Law*, Edward Elgar, Cheltenham

Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2044797 (accessed: 19th January 2026).

Davies, H., McKernan, B. and Sabbagh, D., 2023, ‘“The Gospel”: how Israel uses AI to select bombing targets in Gaza’, *The Guardian*, Fri 1 Dec 2023 10.03 GMT:

<https://www.theguardian.com/world/2023/dec/01/the-gospel-how-israel-uses-ai-to-select-bombing-targets> (accessed: 19th January 2026).

Beauvoir, S. de, 1956, *The Second Sex*, H. M. Parshley (trans., ed.), Jonathan Cape: Thirty Bedford Square London

de las Heras Ballell, T., 2010, *Introduction to Spanish Private Law: Facing the Social and Economic Challenges* Routledge-Cavendish

de Vries, B., 2023, *Individual Criminal Responsibility for Autonomous Weapons Systems in International Criminal Law*, Leiden, The Netherlands: Brill Nijhoff. Available online: <https://doi.org/10.1163/9789004524316> (accessed: 19th January 2026).

Deakin, S., 2003, ‘“Enterprise-Risk”: The Juridical Nature of the Firm Revisited’, *Industrial law journal (London)* 32 (2), 97–114.

DeGrazia, D., 2008, ‘Moral Status As a Matter of Degree?’, *The Southern Journal of Philosophy* Vol. XLVI: 181-198

Deep Learning Glossary, Deep AI. Available online: <https://deepai.org/machine-learning-glossary-and-terms/deep-learning> (accessed: 19th January 2026).

Delvaux, M. 2017, Report with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)), *Committee on Legal Affairs*.

Demirtas, H., 2025, 'AI responsibility gap: not new, inevitable, unproblematic', *Ethics Inf Technol* **27**, 7. Available online: <https://doi.org/10.1007/s10676-024-09814-1> (accessed: 19th January 2026).

Dempsey, M. M., 2023, 'Blame and Punishment: The Difference Duty Makes', in M. M. Dempsey and F. Tanguay-Renaud (eds), *From Morality to Law and Back Again: A Liber Amicorum for John Gardner* (Oxford, 2023; online edn, Oxford Academic, 24 Aug. 2023). Available online: <https://doi.org/10.1093/oso/9780198860594.003.0008> (accessed: 20th January 2026).

Dennett, D., 1987, *The Intentional Stance*, MIT Press

———, 2009, 'Intentional systems theory', in A. Beckermann, B. McLaughlin, & S. Walter (eds.), *The Oxford Handbook of Philosophy of Mind*, Oxford University Press: 339-350

Deprez, F., 2025, 'How AI guided Ukraine's drones to hit Russian airfields', *FT*, 11 June 2025. Available online: <https://www.ft.com/content/ccd83e2a-521f-4e35-a5f0-2ec1ef63749e> (accessed: 19th January 2026).

Devlin, K., 2018, *Turned on: Science, sex and robots*, London: Bloomsbury Publishing

———, 2020, 'Meet Samantha, The Sex Robot That Gives You Life Advice', *Refinery 29*. Available online: <https://www.refinery29.com/en-gb/sex-robots-kate-devlin> (accessed: 19th January 2026).

Devlin, P., [1965]1968, *The Enforcement of Morals*, Oxford: Oxford University Press

- Dewey, J., 1926, 'The Historic Background of Corporate Legal Personality', 35 *Yale Law Journal*
- Dignum, V., 2020, 'Responsibility and Artificial Intelligence', in M.D. Dubber (ed.) *The Oxford Handbook of Ethics in AI*, Oxford.
- Donner, W., 2008, 'Autonomy, Tradition, and the Enforcement of Morality', in C. L. Ten (ed.) *Mill's 'On Liberty': A Critical Guide*, New York: Cambridge University Press
- Draft Articles on the Responsibility of States for Internationally Wrongful Acts, 2001.
Available online: https://legal.un.org/ilc/texts/instruments/english/draft_articles/9_6_2001.pdf
(accessed: 20th January 2026).
- Duff, A., 2001, *Punishment, Communication, and Community* (Oxford: OUP, 2001)
- , 2009, 'Answering for Crime: Responsibility and Liability in the Criminal Law', *Legal Theory Today*, Oxford: Hart.
- , 2014, 'Towards a Modest Legal Moralism', *Criminal Law and Philosophy* 8: 217-235
- , 2018, *The Realm of Criminal Law*, OUP
- Duffy, B., 2003, *Anthropomorphism and the social robot*, *ROBOTICS AND AUTONOMOUS SYSTEMS* 42: 179-83
- Dunham, J. and Lawford-Smith, H., 2017, 'Offsetting Race Privilege', *Journal of Ethics & Social Philosophy* 11:1-22.
- Dworkin, R., 1978, 'Liberalism,' in *Public and Private Morality*, S. Hampshire (ed.), Cambridge: Cambridge University Press, pp. 113–43.
- , 1986, *Law's Empire*, Belknap Press

——, 1995, 'Foundations of Liberal Equality,' in *Equal Freedom*, S. Darwall (ed.), Ann Arbor: University of Michigan Press.

——, 2011, *Justice for Hedgehogs*, Cambridge, Mass.: Harvard University Press

Dyzenhaus, D., 1992, 'John Stuart Mill and the Harm of Pornography', *Ethics* 102: 534-551

Dwyer, S. (ed.), 1995, *The Problem of Pornography*, Belmont CA: Wadsworth

Dylan, B., 1963, 'Masters of War', *Biograph*. Available online: <https://genius.com/Bob-dylan-masters-of-war-lyrics> (accessed: 21st January 2026).

——, 1964, 'The times they are a-changin''. Available online: <https://genius.com/Bob-dylan-the-times-they-are-a-changin-lyrics> (accessed: 21st January 2026).

E&T editorial staff, 2017, 'Social implications of sex robots scrutinised in new report', July 5 2017.

Ebels-Duggan, K., 2010, 'The Beginning of Community: Politics in the Face of Disagreement,' *The Philosophical Quarterly*, 60(283): 50–71

Ecuador Constitution, Available online: https://www.constituteproject.org/constitution/Ecuador_2021?lang=en (accessed: 19th January 2026).

Edgington, D., 2011, 'Causation First: Why Causation is Prior to Counterfactuals' in C. Hoerl, T. McCormack, and S. Beck (eds.), *Understanding Counterfactuals, Understanding Causation: Issues in Philosophy and Psychology*: 230-241, Oxford: Oxford University Press

Edmundson, W. A., 2004, *An Introduction to Rights*, Cambridge Introductions to Philosophy and Law), second edition, Cambridge/New York: Cambridge University Press; second edition, 2012

Eggert, L., 2022a, 'LAW AND MORALITY IN HUMANITARIAN INTERVENTION', *Legal Theory* 28, no. 4: 298–324.

——, 2022b, 'Handle With Care: Autonomous Weapons and Why the Laws of War Are Not Enough', J. Simons (ed.), Cambridge, MA: Harvard Kennedy School.

——, 2023, 'Autonomised harming', *Philosophical Studies* (2025) 182:1-24. Available online: <https://doi.org/10.1007/s11098-023-01990-y> (accessed: 19th January 2026).

——, 2024, 'Rethinking 'Meaningful Human Control'', in J. M. Schraagen (ed.) *Responsible Use of AI in Military Systems*, Chapman and Hall/CRC:213-231.

Eidenmuller, H. and Wagner, G., 2021, *Law by Algorithm*, Mohr Siebeck

Einsiedel, E., 1992, 'The Experimental Research Evidence; Effects of Pornography on the 'Average Individual'', in C. Itzin (ed.) *Pornography: Women , Violence and Civil Liberties*, Oxford: Oxford University Press: 248-283

Ekelhof, M., 2018, 'Autonomous weapons: Operationalizing meaningful human control', *Humanitarian Law & Policy*, 15 August 2018. Available online: <https://blogs.icrc.org/law-and-policy/2018/08/15/autonomous-weapons-operationalizing-meaningful-human-control/> (accessed: 19th January 2026).

Eliot, T. S., 1921, review of *Metaphysical Lyrics and Poems of the Seventeenth Century: Donne to Butler*. Selected and edited, with an Essay, by Herbert J. C. Grierson (Oxford: Clarendon Press. London; Milford) in the *Times Literary Supplement*, October 1921

Epstein, R., 1985, 'Products Liability as an Insurance Market', *J. Legal Stud.* 14:645

Eskens, R., 2017, 'Is Sex With Robots Rape?', *Journal of Practical Ethics*: 62-76

EU AI Act Regulation 2024/1689 ('EU AI Act')

European Product Liability Directive (85/374/EEC). Available online: <https://eur-lex.europa.eu/eli/dir/1985/374/oj/eng> (accessed: 20th January 2026).

EU Product Liability Directive (2024/2853). Available online: <https://eur-lex.europa.eu/eli/dir/2024/2853/oj/eng> (accessed: 20th January 2026).

European Commission, 2025, ‘COMMUNICATION FROM THE COMMISSION Commission Guidelines on the definition of an artificial intelligence system established by Regulation (EU) 2024/1689 (AI Act)’ (29th July 2025). English translation available online: <https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-ai-system-definition-facilitate-first-ai-acts-rules-application> (accessed: 22nd December 2025); <https://ec.europa.eu/newsroom/dae/redirection/document/112455> (direct download link) (accessed: 24th January 2026).

European Economic and Social Committee, 2017, ‘Opinion on ‘Artificial Intelligence—The Consequences of Artificial Intelligence on the (Digital) Single Market, Production, Consumption, Employment and Society’ (Own-Initiative Opinion) OJ C288/1’

European Parliament Resolution 2020/2014[2021], European Parliament resolution of 20 October 2020 with recommendations to the Commission on a civil liability regime for artificial intelligence (2020/2014(INL)). Available online: <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020IP0276> (accessed: 19th January 2026).

Ezra (10:4), *New American Standard Bible*, 1960, 1971, 1977, 1995, 2020, The Lockman Foundation. Available online: <https://www.biblegateway.com/passage/?search=Ezra%2010%3A3-5&version=NASB> (accessed: 25th January 2026).

Fara, M., 2008, ‘Masked Abilities and Compatibilism’, *Mind* 117(468): 843-865

Fatima, Z., 2025, 'Football clubs should pay towards £70m policing cost, Met chief says', *BBC News* (6 July 2025). Available online: <https://www.bbc.co.uk/news/articles/cqx2338z971o> (accessed: 14th January 2026).

Feinberg, J., 1965, 'The Expressive function of punishment', *The Monist*, 49(3):397–423.

——, 1970, *Doing & deserving; essays in the theory of responsibility*. Princeton University Press.

——, 1984, *Harm to others*, New York: Oxford University Press

Feldman, M., Friedler, S., Moeller, J., Scheidegger, C., Venkatasubramian, S., 2015, 'Certifying and removing disparate impact', 2015 ACM SIGKDD Conference on Knowledge Discovery and Data Mining. Available online: <https://arxiv.org/abs/1412.3756> (accessed: 19th January 2026).

Ferguson, C. J., & Hartley, R. D., 2022, 'Pornography and Sexual Aggression: Can Meta-Analysis Find a Link?', *Trauma, Violence, & Abuse*, 23(1): 278–287

Fernández, V., Matteo, A., de Mendonça, J. T., Tighe, P., 2025, 'Antiqua et Nova: Note on the Relationship Between Artificial Intelligence and Human Intelligence', DICASTERY FOR THE DOCTRINE OF THE FAITH, DICASTERY FOR CULTURE AND EDUCATION (14 January 2025). Available online here: https://www.vatican.va/roman_curia/congregations/cfaith/documents/rc_ddf_doc_20250128_antiqua-et-nova_en.html (accessed: 24th January 2026).

Fine, K., 1975, 'Review of Counterfactuals', *Mind* 84:451-458.

Finnis, J., 1981, *Natural law and natural rights*, 1st ed., Oxford: Oxford University Press

——, 1997a, 'The Good of Marriage and the Morality of Sexual Relations: Some Philosophical and Historical Observations', *American Journal of Jurisprudence* 42

- , 1997b, ‘Law, Morality, and ‘Sexual Orientation’’, in J. Corvino (ed.), *Same Sex: Debating the Ethics, Science, and Culture of Homosexuality*, Rowman and Littlefield
- , 2008, ‘Marriage: A Basic and Exigent Good’, *The Monist*, Vol. 91, Nos. 3 & 4: 388-406
- , 2011a, *Natural law and natural rights*, 2nd ed. Oxford: Oxford University Press
- , 2011b, ‘The Priority of Persons’, in J. Finnis, *Intention and Identity*, OUP
- , 2013, ‘The Priority of Persons Revisited’, *The American Journal of Jurisprudence* 58(1): 45-62
- Fischer, J.M., 2004, ‘Responsibility and Manipulation’, *The Journal of Ethics* 8, 145–177.
Available online: <https://doi.org/10.1023/B:JOET.0000018773.97209.84> (accessed: 20th January 2026).
- Fischer, J. M. and Ravizza, M., 1998, *Responsibility and Control: A Theory of Moral Responsibility*, New York: Cambridge University Press.
- Fischer, J. M., & Tognazzini, N. A., 2009, ‘The truth about tracing’, *Noûs*, 43(3):531–556
- Fitzgerald, P. J., 1966, *Salmond on Jurisprudence*, twelfth edition, London: Sweet & Maxwell
- Fleming, J., 1998, *The Law of Torts*, (9th ed.), LBC.
- Forrest, K. B., 2021, ‘When Machines Can Be Judge, Jury, and Executioner: Justice in the Age of Artificial Intelligence’, *World Scientific Publishing Co. Pte. Ltd.*
- Frank, L. and Nyholm, S., 2017, ‘Robot sex and consent: Is consent to sex between a robot and a human conceivable, possible, and desirable?’, *Artificial Intelligence Law* 25: 305-323
- Frankfurt, H., 1969, ‘Alternate Possibilities and Moral Responsibility’, *Journal of Philosophy* 66(23): 829-839

French Civil Code. Available online:
https://www.legifrance.gouv.fr/codes/texte_lc/LEGITEXT000006070721/ (accessed: 20th
January 2026).

Freund, E., [1897]2000, *The Legal Nature of Corporations*, Batoche Books

Fricker, M., 2003, 'Epistemic Injustice and a Role for Virtue in the Politics of Knowing', *Metaphilosophy*, 34(1–2): 154–173

——, 2007, *Epistemic Injustice: Power and the Ethics of Knowing*, Oxford: Oxford University Press

Friedman, B., Kahn, P., Hagman, J., 2003, 'Hardware Companions? – What Online AIBO Discussion Forums Reveal about the Human-Robotic Relationship', *PROCEEDINGS OF THE SIGCHI CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS*: 273–80

Fritz, N. and Paul, B., 2017, 'From Orgasms to Spanking: A Content Analysis of the Agentic and Objectifying Sexual Scripts in Feminist, for Women, and Mainstream Pornography', *Sex Roles* 77: 639–52

Frowe, H., 2023, *The ethics of war and peace: an introduction*, New York, NY: Routledge, Taylor & Francis Group.

Frydrych, D., 2018, 'State-of-the-Art: The Theories of Rights Debate' (February 13, 2018), *Jurisprudence: An International Journal of Legal and Political Thought* 9(3): 566-588

Gagnon, J. H. and Simon, W., 1973, *Sexual Conduct: The Social Sources of Human Sexuality*, Aldine

Gaita, R., 1999, *A Common Humanity: Thinking About Love & Truth & Justice*, Text Publishing, Melbourne

Galaitis, S. E., Hendren, C. O., Trump, B. and Linkov, I., 2019, 'Sex Robots—A Harbinger for Emerging AI Risk', *Frontiers of Artificial Intelligence*

Galliot, J., 2015, *Military Robots: Mapping the Moral Landscape*, Surrey UK: Ashgate

Galliot, J., D. MacIntosh and J.D. Ohlin (eds.), 2021, *Lethal Autonomous Weapons: Re-Examining the Law and Ethics of Robotic Warfare*, Oxford University Press

Gardner, J., 2005, 'Wrongs and Faults', in *Appraising Strict Liability*, A. Simester (ed.): 51–80, Oxford: Oxford University Press.

———, 2021, 'Why Blame?' in I. Solanke (ed), *On Crime, Society, and Responsibility in the Work of Nicola Lacey* (OUP 2021) 85.

Gardner, J., and Shute, S., 2000, 'The Wrongness of Rape', in J. Horder (eds.) *Oxford Essays in Jurisprudence*, Oxford University Press

Garreau, J., 2007, 'Bots on The Ground in the Field of Battle (Or Even Above It), Robots Are a Soldier's Best Friend', *WASHINGTON POST* (May 6, 2007)

Gaus, G., 1996, *Justificatory Liberalism: An Essay on Epistemology and Political Theory*, Oxford: Oxford University Press

———, 2009, 'The Place of Religious Belief in Public Reason Liberalism,' in *Multiculturalism and Moral Conflict*, M. Dimova-Cookson and P. Stirk (eds.), New York: Routledge, pp. 19–37

———, 2011, *The Order of Public Reason: A Theory of Freedom and Morality in a Diverse and Bounded World*, Cambridge: Cambridge University Press

———, 2012, 'Sectarianism Without Perfection? Quong's Political Liberalism,' *Philosophy and Public Issues*, 2(1): 7–15

———, 2016, *The Tyranny of the Ideal*, Princeton, NJ: Princeton University Press

——, 2017, ‘Self-Organizing Moral Systems,’ *Politics, Philosophy, & Economics*, first published online 29 August 2017. doi: 10.1177/1470594X17719425

Gaus, G, and K. Vallier, 2009, ‘The Roles of Religious Conviction in a Publicly Justified Polity: The Implications of Convergence, Asymmetry, and Political Institutions,’ *Philosophy & Social Criticism*, 35(1): 51–76

Geneva Convention, Additional Protocol I, Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of International Armed Conflicts (Protocol I) art. 36, June 8, 1977, 1125 U.N.T.S. 3

George, R., 2004, ‘What’s Sex Got To Do With It? Marriage, Morality, and Rationality’, *The American Journal of Jurisprudence* 49(1): 63-85

German Civil Code, *Bürgerliches Gesetzbuch*, BGBl, Available online: <https://www.gesetze-im-internet.de/bgb/BGB.pdf> (accessed: 19th January 2026).

Gersen, J. S., 2019, ‘Sex Lex Machina’, *Columbia Law Review* Vol. 199, No. 7: 1793-1810

Giliker, P., 2016, ‘Vicarious liability in the UK Supreme Court’, in *UK Supreme Court Yearbook* Vol. 7:152-166

— (ed.), 2022, *Vicarious Liability in the Common Law World*, *Hart Studies in Private Law*, Bloomsbury Publishing

Ginet, C., 1990, *On Action*, Cambridge: Cambridge University Press

Gips, J., 1995, ‘Towards the Ethical Robot’, Second International Workshop on Human & Machine Cognition, Perdido Key, Florida, 1991, published in *Android Epistemology*, K. Ford, C. Glymour, P. Hayes (eds.), MIT Press

Glavaničová, D., Pascucci, M., 2022, 'Vicarious liability: a solution to a problem of AI responsibility?' *Ethics Inf Technol* 24:28 Available online: <https://doi.org/10.1007/s10676-022-09657-8> (accessed: 19th January 2026).

Global Security, 'Military: Samsung Techwin SGR-A1 Sentry Guard Robot'. Available online: <https://www.globalsecurity.org/military/world/rok/sgr-a1.htm> (accessed: 5th January 2026).

Goldberg, S., 2010, *Relying on Others: An Essay in Epistemology*, New York: Oxford University Press

———, 2013, 'Anonymous assertions', *Episteme* 10(2): 134–151

———, 2015, *Assertion: On the Philosophical Significance of Assertoric Speech*, Oxford: Oxford University Press

Goldblatt, B., 2012, 'Virtual Child Pornography: The Children Aren't Real, But the Dangers Are; Why the Ashcroft Court Got it Wrong', *Student Works* 41. Available online: https://scholarship.shu.edu/student_scholarship/41 (accessed 4th January 2026).

Golding, M. P. and W. A. Edmundson (eds.), 2005, *The Blackwell Guide to the Philosophy of Law and Legal Theory*, Oxford: Blackwell Publishing

Goodin, R., 2013, 'Disgorging the Fruits of Historical Wrongdoing', *American Political Science Review* 107:478-491.

Gosseries, A., 2004, 'Historical emissions and free-riding', *Ethical Perspectives* 11(1): 38–62.

Gray, J. C., 1996, *Mill on Liberty: A Defence*, 2nd ed., New York: Routledge

———, 1997, *The Nature and Sources of the Law*, D. Campbell and P. Thomas (eds.), Ashgate

Green, L., 2013, 'Should Law Improve Morality?', *Criminal Law and Philosophy* 7: 473-494

Green, S., 2020, 'Legal Moralism, Overinclusive Offenses, and the Problem of Wrongfulness Conflation', *Criminal Law & Philosophy* 14: 417-430

Greshko, M., 2018, 'Meet Sophia, the Robot That Looks Almost Human', *National Geographic*. Available online: <https://www.nationalgeographic.com/photography/article/sophia-robot-artificial-intelligence-science> (accessed: 19th January 2026).

Grieman, K., 2024, 'Law, Death, and Robots: The Regulation of Artificial Intelligence in High-Risk Civil Applications', Hart Publishing (UK).

Gunkel, D. J., 2012, *The machine question: critical perspectives on AI, robots, and ethics*, MIT Press, Cambridge, MA

Gutiu, S., 2012, 'Sex Robots and the Roboticization of Consent', *We Robot Law Conference Miami*

——, 2016, 'The Roboticization of Consent', in *Robot Law*, Calo, Froomkin and Kerr (eds.), Cheltenham, UK; Edward Elgar Publishing

Habermas, J., 1990, *Moral Consciousness and Communicative Action*, C. Lenhardt and S. W. Nicholsen (trans.), Cambridge, MA: MIT Press

——, 1995, 'Reconciliation Through the Public use of Reason: Remarks on John Rawls's Political Liberalism,' *The Journal of Philosophy*, 92(3): 109–131

——, 1996, *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy*, W. Rehg (trans.), Cambridge, MA: MIT Press

——, 1998, *The Inclusion of the Other: Studies in Political Theory*, C. Cronin and HP. DeGreiff (eds.), Cambridge, MA: MIT Press

- Hallis, F., 1930, *Corporate Personality. A Study in Jurisprudence*, OUP
- Hampton, J., 1984, 'The Moral Education Theory of Punishment', *Philosophy & Public Affairs*, Vol. 13, No. 3: 208-238
- Haydar, B. and Øverland, G., 2014, 'The Normative Implications of Benefiting from Injustice', *Journal of Applied Philosophy* 31:349-362.
- Hänel, H., 2021, 'Who's to Blame? Hermeneutical Misfire, Forward-Looking Responsibility, and Collective Accountability', *Social Epistemology* 35(2): 173-184
- Hanson, F. A., 2009, 'Beyond the skin bag: On the moral responsibility of extended agencies', *Ethics and Information Technology*, 11(1), 91–99.
- Hanson Robotics, 2023, 'Sophia'. Available online: <https://www.hansonrobotics.com/sophia/> (accessed: 19th January 2026).
- Harel, A., 2005, 'Theories of Rights', in Golding and Edmundson 2005: 191–206
- Hart, H. L. A., 1961, *The Concept of Law*, Oxford University Press
- , [1964]1968, *Law, Liberty and Morality*, Oxford: Oxford University Press
- , 1968, *Punishment and responsibility*, Oxford University Press
- , 1982, *Essays on Bentham: Studies in Jurisprudence and Political Theory*, Oxford: Clarendon Press
- , 1983, *Essays in Jurisprudence and Philosophy*, Oxford: Oxford University Press
- Hart, H. L. A. and Honoré, T., 1985, 'Causation in the Law', 2nd edition, Oxford: Oxford University Press
- Hartley, C. and L. Watson, 2010, 'Is a Feminist Political Liberalism Possible?' *Journal Of Ethics & Social Philosophy*, 5(1): 1–21

Haslanger, S., 2005, 'What Are We Talking About? The Semantics and Politics of Social Kinds', *Hypatia*, 20(4): 10–26

Hegel, F., Muhl, C., Wrede, B., Hielscher-Fastabend, M., and Sagerer, G., 2009, *Understanding Social Robots*, The Second International Conferences on Advances in Computer-Human Interactions: 169-74.

Heinze, A., 2020, 'Chapter 1 Bridge over Troubled Water – A Semantic Approach to Purposes and Goals in International Criminal Justice', in *The Politics of International Criminal Law*, Leiden, The Netherlands: Brill | Nijhoff. Available online: https://doi-org.ezproxy-prd.bodleian.ox.ac.uk/10.1163/9789004372498_003 (accessed: 22nd January 2026).

Herman, D., 1988, 'The rape culture', *Culture* 1(10): 45-53

Hern, A., 2017, 'Give robots 'personhood' status, EU committee argues', *The Guardian*

Herzfeld, N., 2017, 'Religious Perspectives on Sex with Robots', in J. Danaher and N. McArthur (eds.) *Robot Sex: Social and Ethical Implications*: 122-136

Heuston, R. E. V., Buckley, R. A., *Salmond and Heuston On The Law of Torts*, Sweet Maxwell Ltd. - London - 21st Edition - 1996).

Heyns, C., 2013, UN Human Rights Council, Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions, Christof Heyns, Lethal Autonomous Robotics, A/HRC/23/47, April 9, 2013. Available online: http://www.ohchr.org/Documents/HRBodies/HRCouncil/RegularSession/Session23/A-HRC-23-47_en.pdf (accessed: 19th January 2026).

Hill, K., 2014, 'Are Child Sex-Robots Inevitable?', *Forbes*

Available online: <https://www.forbes.com/sites/kashmirhill/2014/07/14/are-child-sex-robots-inevitable/?sh=529beb0e460b> (accessed: 19th January 2026)

Himmelreich, J., 2019, 'Responsibility for Killer Robots', *Ethical Theory and Moral Practice* 22: 731-747

Hines, D., 2009, founder of TrueCompanion, cited in TechCrunch, 2010, 'NSFW: True Companion Debuts Sex Robot Roxxxy'. Available online https://techcrunch.com/2010/01/09/nsfw-true-companion-debuts-sex-robot-roxxxy/?guccounter=1&guce_referrer=aHR0cHM6Ly9kdWNrZHVja2dvLmNvbS8&guce_referrer_sig=AQAAANVNI210M_gyuakfa2Ua14z6F-Jq5Y6-GGNgVtit5sAb1szozbrprKWuQcji3dypMnITZBriRuXZmvX7QRUAbp882P7I8jWLblikACnYrBlKelNTV_aM31c27ubi1WmS_j67VaQCCPXBOuAtmnhA5ORCfQgD4UVbonMphcd2v15l (accessed: 19th January 2026).

Horowitz, M. and Scharre, P., 2015, 'An Introduction to Autonomy in Weapon Systems', unpublished working paper. Available online: https://www.files.ethz.ch/isn/188865/Ethical%20Autonomy%20Working%20Paper_021015_v02.pdf (accessed: 19th January 2026).

Hohfeld, W. N., 1909, 'Nature of Stockholders' Individual Liability for Corporation Debts', *Columbia Law Review* 9(4): 285-320

——, 1919, *Fundamental Legal Conceptions*, Walter Wheeler Cook (ed.), New Haven, CT: Yale University Press

Hobart, R. E., 1934, 'Free Will as Involving Indeterminism and Inconceivable Without It', *Mind* 43: 1-27

Horgan, T., 1979, '“Could”, Possible Worlds, and Moral Responsibility', *Southern Journal of Philosophy* 17: 345-358

Horwitz, M., 1985, '*Santa Clara Revisited: The Development of Corporate Theory*', *West Virginia Law Review* 88

HRW (Human Rights Watch), 2012, 'Losing Humanity: The Case against Killer Robots', Human Rights Watch and International Human Rights Clinic at Harvard Law School. Available online: <https://www.hrw.org/report/2012/11/19/losing-humanity/case-against-killer-robots> (accessed: 19th January 2026).

—, 2015, '*Mind the Gap: The Lack of Accountability for Killer Robots*', Human Rights Watch and International Human Rights Clinic at Harvard Law School. Available online: https://www.hrw.org/sites/default/files/reports/arms0415_ForUpload_0.pdf (accessed: 19th January 2026).

—, 2018, 'Heed the Call: A Moral and Legal Imperative to Ban Killer Robots'. Available online: <https://www.hrw.org/report/2018/08/21/heed-call/moral-and-legal-imperative-ban-killer-robots> (accessed: 19th January 2026).

Husak, D., 2007, *Overcriminalization: The limits of the criminal law*, Oxford: Oxford University Press

—, 2010, 'Answering Duff: R. A. Duff's Answering for Crime (Book Review)', *Law and Philosophy* 29: 101–19.

Huxley, A., 2002, *Brave New World*, Available online: <https://plato-philosophy.org/wp-content/uploads/2016/05/BraveNewWorld-1.pdf> (accessed: 19th January 2026).

IAI 'HARPY', 'Anti-Radiation Loitering Munition'. Available online: <https://www.iai.co.il/p/harpy> (accessed: 5th January 2026).

IAMD ‘C-RAM’, *Integrated Air & Missile Defence Centre of Excellence, Counter Rockets, Artillery and Mortars (C-RAM)*. Available online: <https://iamd-coe.org/focus-areas/counter-rockets-artillery-and-mortars-cram/> (accessed: 5th January 2026).

ICRC, 2014, ‘Autonomous weapon systems: Technical, military, legal and humanitarian aspects’, *Expert Meeting Report*, 26-28 March 2014. Available online: https://www.icrc.org/sites/default/files/document/file_list/4221-002-autonomous-weapons-systems-full-report.pdf (accessed: 20th January 2026).

—, 2016, ‘AUTONOMOUS WEAPON SYSTEMS IMPLICATIONS OF INCREASING AUTONOMY IN THE CRITICAL FUNCTIONS OF WEAPONS’, *Expert Meeting Report*, 15-16 March 2016. Available online: https://peterasaro.org/writing/4283_002_Autonomus-Weapon-Systems_WEB.pdf (accessed: 19th January 2026).

—, 2019, ‘Autonomy, artificial intelligence and robotics: Technical aspects of human control’, *International Committee of the Red Cross* (Geneva, August 2019). Available online: https://www.icrc.org/sites/default/files/document/file_list/autonomy_artificial_intelligence_and_robotics.pdf (accessed online: 14th January 2026).

ICRC Commentary, 1987, ‘Commentary on the Additional Protocols of 8 June 1977 to the Geneva Conventions of 12 August 1949’, (Y. Sandoz, C. Swinarski & B. Zimmerman eds.)

ICRC Rule 156. Definition of War Crimes, INT’L COMMITTEE RED CROSS CUSTOMARY INT’L HUMANITARIAN L. DATABASE. Available online: <https://perma.cc/E3P3-9LHM> (accessed: 20th January 2026).

ICTR Statute, Statute of the International Criminal Tribunal for Rwanda, Nov. 8, 1994, 33 I.L.M. 1598

ICTY Statute, Updates Statute of the International Criminal Tribunal for the Former Yugoslavia, May 25, 1993, 32 I.L.M.

Irwin, T., 1980, 'Reason and Responsibility in Aristotle', in A. O'Rorty (ed.) *Essays on Aristotle's Ethics*, Major Thinkers Series, Vol. 2, Berkeley: University of California Press

Ishiguro, K., 2021, *Klara and the Sun*, Faber & Faber

Italian Civil Code, Available online: https://www.trans-lex.org/601300/_italian-codice-civile/ (accessed: 19th January 2026).

Jain, N., 2016, 'Autonomous Weapons Systems: New Frameworks for Individual Responsibility', in N. Bhuta et al. (eds), *Autonomous Weapons Systems: Law, Ethics, Policy* (Cambridge University Press, 2016) 303.

JCRAS Report, 2016, 'Joint Concept for Robotic and Autonomous Systems', October 19, 2016.

Jecker, N., 2020, 'Nothing to be ashamed of: sex robots for older adults with disabilities', *Journal of Medical Ethics* 47: 26-32

Jenkins, H., 2012, 'The war between effects and meaning: rethinking the video game violence debate', in D. Buckingham, R. Willett (eds.) *Digital generations: children, young people, and new media*, Lawrence Erlbaum Associates, Mahwah: 19-32

Jones, P., 1994, *Rights*, New York: St. Martin's Press

——, 1999, 'Human Rights, Group Rights, and Peoples' Rights', *Human Rights Quarterly* 21(1): 80-107

- Kahn, P., Kanda, T., Ishiguro, H., Freier, N., Severson, R., Gill, B., Ruckert, J., Shen, S., 2012, “‘Robovie, you'll have to go into the closet now’: Children's social and moral relationships with a humanoid robot’, *DEVELOPMENTAL PSYCHOLOGY*, 48(2): 303-14
- Kaiserman, A., 2016, ‘Causal Contribution’, *Proceedings of the Aristotelian Society* 116(3):387-394
- , 2017a, ‘Partial Liability’, *Legal Theory* 23(1):1-26
- , 2017b, ‘Necessary Connections in Context’, *Erkenntnis* 82(1):45-64
- , 2020, ‘Reasons-sensitivity and degrees of free will’, *Philosophical and Phenomenological Research*: 1-23
- , 2021a, ‘Against Accomplice Liability’, in J. Gardner, L. Green, and B. Leiter (eds.), *Oxford Studies in Philosophy of Law*: Vol.4, Oxford: OUP.
- , 2021b, ‘Responsibility and the ‘Pie Fallacy’’, *Philosophical Studies* 178:3597-3616. Available online: <https://doi.org/10.1007/s11098-021-01616-1> (accessed: 19th January 2026).
- Kaiserman, A., Beebe, H., 2020, ‘Causal Contribution in War’, in *Journal of Applied Philosophy* 37(3): 364-365
- Kamm, F., 2002, KAMM, *Rights*, in OXFORD HANDBOOK OF JURISPRUDENCE AND PHILOSOPHY OF LAW 484, S. J. Schapiro & J. Coleman, (eds.)
- , 2007, *Intricate Ethics: Rights, Responsibilities, and Permissible Harm*, OUP.
- Karner, E., 2019, ‘Liability for Robotics: Current Rules, Challenges, and the Need for Innovative Concepts’ in S. Lohsse, R. Schulze and D. Staudenmayer (eds.), *Liability for Artificial Intelligence and the Internet of Things*, Nomos 2019

Karnow, C., 1996, 'Liability for distributed Artificial Intelligences', *Berkeley Technology Law Journal*, Vol.11, No.1:147–204.

Keating, G., 2017, 'Products Liability as Enterprise Liability', *J. Tort Law* 10(1): 41-97.

—, 2023, *Reasonableness and Risk: Right and Responsibility in the Law of Torts*, OUP.

Kelsen, H., 1919, 'Zur Theorie der juristischen Fiktionen. Mit besonderer Berücksichtigung von Vaihingers Philosophie des Als Ob', *Annalen der Philosophie* 1: 630-658

—, 1923, *Hauptprobleme der Staatsrechtslehre: entwickelt aus der Lehre vom Rechtssatze* (2nd edn.), JCB

Mohr

—, 1925, *Allgemeine Staatslehre*, Julius Springer

—, 1945, *General Theory of Law and State*, Harvard University Press

—, 1991, *General Theory of Norms*, M. Harntney (trans.), Clarendon Press

—, 1992, *Introduction to the Problems of Legal Theory*, B.L. Paulson and S.L. Paulson (trans.), Oxford: Clarendon Press

—, 2005, *Pure Theory of Law*, M. Knight (tr., ed.), The Lawbook Exchange, Ltd.

—, 2006, *General Theory of Law and State*, Transaction Publishers

—, 2015, 'On the Theory of Juridic Fictions. With Special Consideration of Vaihinger's Philosophy of the As-If', in M. Del Mar and W. Twining (eds.), C. Kletzer (trans.), *Legal Fictions in Theory and Practice* 110, Springer International Publishing

Khoury, A. C., 2012, 'Responsibility, Tracing, and Consequences', *Canadian Journal of Philosophy*, 42(3–4): 187–207.

Khritankov, A., 2021, 'Hidden Feedback Loops in Machine Learning Systems: A Simulation Model and Preliminary Results' in D. Winkler, Biffl, S., Mendez, D., Wimmer, M., Bergsmann, J. (eds) *Software Quality: Future Perspectives on Software Engineering Quality*, SWQD 2021. Lecture Notes in Business Information Processing, vol 404. Springer, Cham. Available online: https://link.springer.com/chapter/10.1007/978-3-030-65854-0_5 (accessed: 4th January 2026).

——, 2023, 'Positive feedback loops lead to concept drift in machine learning systems', *Applied Intelligence* 53: 22648–22666.

Kiener, M., 2022a, 'Are We Heading Towards a Post-Responsibility Era? Artificial Intelligence and the Future of Morality', *Practical Ethics in the News*

——, 2022b, 'Can we Bridge AI's responsibility gap at Will?', *Ethical Theory and Moral Practice* 25:575–593. Available online: <https://doi.org/10.1007/s10677-022-10313-9> (accessed: 19th January 2026).

——, 2023, 'Varieties of answerability', *The Routledge Handbook of Philosophy of Responsibility*. Available online: <https://www.taylorfrancis.com/chapters/edit/10.4324/9781003282242-23/varieties-answerability-maximilian-kiener> (accessed: 19th January 2026).

——, 2024, 'Strict Moral Answerability', *Ethics* 134 (3), 360-386

——, 2025, 'AI and Responsibility: No Gap, but Abundance', *Journal of Applied Philosophy*, 42 (1), 357-374.

King, M., 2009, 'The Problem with Negligence', *Social Theory and Practice*, 35(4): 577–595.

——, 2014, 'Traction without Tracing: A (Partial) Solution for Control-Based Accounts of Moral Responsibility', *European Journal of Philosophy*, 22(3): 463–482

Kirkpatrick, K., 2017 (2017-01-23), 'It's not the algorithm, it's the data', *Communications of the ACM* 60(2): 21–23

Kirste, S., 2000, 'Verlust und Wiederaneignung der Mitte—zur juristischen Konstruktion der Rechtsperson', *Evangelische Theologie* 60(25)

Kleeman, J., 2017, 'The race to build the world's first sex robot', *The Guardian*, 27 April 2017. Available online: <https://www.theguardian.com/technology/2017/apr/27/race-to-build-world-first-sex-robot> (accessed: 19th January 2026).

———, 2020, *Sex Robots & Vegan Meat*, Picador

Klosko, G., 2003, 'Reasonable rejection and neutrality in justification', in *Perfectionism and neutrality: Essays in liberal theory*, S. Wall and G. Klosko (eds.), Oxford: Rowman & Littlefield: 167-189

Kment, B., 2010, 'Causation: Determination and Difference-Making', *Noûs* 44: 80-111

Knight, H., 2014, *How Humans Respond to Robots: Building Public Policy through Good Design*, *BROOKINGS REPORT*

Knight, W., 2017, 'The Dark Secret at the Heart of AI (April 11th)', *MIT Technology Review*.

Köhler, S., Roughley, N. and Sauer, H., 2017, 'Technologically blurred accountability?' in C. Ulbert *et al.* (eds), *Moral Agency and the Politics of Responsibility*, Routledge.

Kolb, M., 2012, *SOLDIER AND ROBOT INTERACTION IN COMBAT ENVIRONMENTS*, Dissertation, University of Oklahoma (2012).

Koops, B.-J., Hildebrandt, M., & Jaquet-Chifelle, D.-O., 2010, 'Bridging the accountability gap: Rights for new entities in the information society' *Minnesota Journal of Law, Science and Technology*

Korosec, K., 2015, 'Volvo CEO: we will Accept All Liability when our Cars Are in Autonomous Mode', *Fortune*, 7 October 2015.

Krahé, B., Temkin, J. and Bieneck, S., 2007, 'Schema-driven information processing in judgements about rape', *Appl. Cognit. Psychol.* 21: 601-619

Kramer, M. H., 1998, 'Rights Without Trimmings', in Kramer, Simmonds, and Steiner 1998: 7–112

——, 2001, 'Getting Rights Right', in *Rights, Wrongs and Responsibilities*, Matthew H. Kramer (ed.), London: Palgrave Macmillan UK, 28–95. doi:10.1057/9780230523630_2

——, 2010, 'Refining the Interest Theory of Rights', *The American Journal of Jurisprudence*, 55(1): 31–39. doi:10.1093/ajj/55.1.31

——, 2013, 'Some Doubts about Alternatives to the Interest Theory of Rights', *Ethics*, 123(2): 245–263. doi:10.1086/668705

——, 2017, 'In Defense of the Interest Theory of Right-Holding: Rejoinders to Leif Wenar on Rights', in McBride 2017: 49–84 (ch. 3)

Kramer, M. H., N. E. Simmonds, and H. Steiner, 1998, *A Debate over Rights: Philosophical Enquiries*, Oxford: Clarendon Press

——, 2001, 'Getting Rights Right' in M. Kramer (ed.) *Rights, Wrongs and Responsibilities*, Palgrave

Kraska, J., 2021, 'Command Accountability for AI Weapons Systems in the Law of Armed Conflict', 97 *International Law Studies* 407.

Krishnan, A., 2009, *Killer Robots: Legality and Ethicality of Autonomous Weapons*, Routledge.

Kurki, V., 2019, *A Theory of Legal Personhood*, Oxford University Press

- Lackey, J., 2006, 'Learning from words', *Philosophy and Phenomenological Research* 73(1): 77–101
- , 2008, *Learning From Words: Testimony as a Source of Knowledge*, Oxford: Oxford University Press
- Lamond, G., 2007, 'What is a crime?', *Oxford Journal of Legal Studies* 27(4): 609-632
- Lancaster, K., 2021, 'Non-consensual personified sexbots: an intrinsic wrong', *Ethics and Information Technology* 23: 589-600
- Langton, R., 1993, 'Speech Acts and Unspeakable Acts', *Philosophy & Public Affairs*, 22(4): 293-330
- , 2009, *Sexual Solipsism: Philosophical Essays on Pornography and Objectification*, Oxford
- Larenz, K., 1983, *Allgemeiner Teil des deutschen Bürgerlichen Rechts. Ein Lehrbuch*, C. H. Beck'sche Verlagsbuchhandlung
- Larmore, C., 1987, *Patterns of Moral Complexity*, Cambridge: Cambridge University Press
- , 1999, 'The Moral Basis of Political Liberalism', *The Journal of Philosophy* Vol. 96, No. 12: 599-625
- Laukyte, M., 2014, 'Artificial Agents: Some consequences of a few capacities', in J. Seibt *et al.* (eds.) *Sociable Robots and the Future of Social Relations*: 115-122, IOS Press
- , 2017, 'Artificial agents among us: Should we recognise them as agents proper?', *Ethics and Information Technology*, 19(1): 1-17
- Leland, R.J., 2019, 'Civic Friendship, Public Reason,' *Philosophy & Public Affairs*, 47(1): 72–103

- Leland, R.J. and H. van Wietmarschen, 2012, 'Reasonableness, Intellectual Modesty, and Reciprocity in Political Justification,' *Ethics*, 122(4): 721–747
- , 2017, 'Political Liberalism and Political Community,' *Journal of Moral Philosophy*, 14(2): 142–167
- Levinas, E., 1989, 'Ethics as first philosophy', in S. Hand (ed.), *The Levinas Reader*, Basil Blackwell, Oxford: 75-87
- Levy, D., 2007, *Love and Sex with Robots*, Harper Perennial, New York
- Lin, P., Bekey, G. and Abney, K., 2008, 'Autonomous Military Robotics: Risk, Ethics, and Design', US Department of Navy, Office of Naval Research. Available online: https://ethics.calpoly.edu/ONR_report.pdf (accessed: 19th January 2026).
- Lingis, A., 1994, *The Community of Those Who Have Nothing in Common*, Indiana University Press, Bloomington and Indianapolis
- List, C., 2021, 'Group Agency and Artificial Intelligence', *Philosophy & Technology* 34: 1213-1242
- List, C. and Pettit, P., 2011, *Group Agency: The possibility, design, and status of corporate agents*, OUP
- Lister, A., 2010, 'Public Justification and the Limits of State Action,' *Politics, Philosophy, & Economics*, 9(2): 151–175.
- , 2013, *Public Reason and Political Community*, London: Bloomsbury.
- , 2017, 'Public Reason and Reciprocity,' *Journal of Political Philosophy*, 25(2): 155–172
- Lloyd, S., 1998, 'Toward a Liberal Theory of Sexual Equality', *Journal of Contemporary Legal Issues* 9: 203-224

- Lu, J., Liu, A., Dong, F., Gu, F., Gama, J. and Zhang, G., 2019, 'Learning under Concept Drift: A Review', in *IEEE Transactions on Knowledge and Data Engineering*, vol. 31, no. 12: 2346-2363.
- Luck, M., 2009, 'The Gamer's Dilemma', *Ethics and Information Technology* 11(1)
- Luck, M. and Ellerby, N., 2013, 'Has Bartel Resolved the Gamer's Dilemma?', *Ethics and Information Technology* 15(3)
- Macedo, S., 1995, 'Homosexuality and the Conservative Mind', *Georgetown Law Journal* 84
- MacCormick, N., 1977, 'Rights in Legislation', in *Law, Morality and Society: Essays in Honour of H.L.A. Hart*, P.M.S. Hacker, and Joseph Raz (eds), Oxford: Clarendon Press: 189–209
- , [1988]2007, *Institutions of Law: An Essay in Legal Theory*, OUP: 77-99
- Machen, A., 1911, 'Corporate Personality', *Harvard Law Review* 24(253)
- MacKinnon, C., 1987, 'Not a Moral Issue' and 'Francis Biddle's Sister: Pornography, Civil Rights and Speech,' in *Feminism Unmodified*, Cambridge, MA: Harvard University Press
- , 1989, 'Rape: On Coercion and Consent', *Toward a Feminist Theory of the State*, Harvard University Press
- , 1993, *Only Words*, Cambridge: Harvard University Press
- , 2006, 'Defining Rape Internationally: A Comment on Akayesu', *Columbia Journal of Transnational Law* 44(3): 940-958
- , 2016, 'Rape Redefined', *Harvard Law & Policy Review* 10(2): 431-477

- MacWilliam, S., 2023, 'The Voice of the 'Sex Robot': From Peep-Show Bucket to Willing Victim—The Terrorism of Women's Speech' in K. Richardson and C. Odland (eds.) *Man-Made Women*, Palgrave Macmillan
- Mains, B., 2010, 'Virtual Child Pornography, Pandering, and the First Amendment: How Developments in Technology and Shifting First Amendment Jurisprudence Have Affected the Criminalization of Child Pornography', 37 HASTINGS CONST. L. Q. 809.
- Mantouvalou, V. and Wolff, J. (eds.), 2024, *Structural Injustice and the Law*, UCL Press.
- Marchant, G.E. *et al.*, 2011, 'International governance of autonomous military robots' XII *The Columbia Science and Technology Law Review* 288.
- Margulies, P., 2017, 'Making Autonomous Weapons Accountable: Command Responsibility for Computer-Guided Lethal Force in Armed Conflicts', in J. Ohlin (ed.), *Research Handbook on Remote Warfare* (Edward Elgar Publishing, 2017) 405.
- Marino, D., & Tamburrini, G., 2006, 'Learning robots and human responsibility', *International Review of Information Ethics*, 6(12): 46–51.
- Marra, W. and McNeil, S., 2012, 'UNDERSTANDING 'THE LOOP': REGULATING THE NEXT GENERATION OF WAR MACHINES', *Harvard Journal of Law & Public Policy* Vol.36(3): 1140-1185.
- Matthias, A., 2004, 'The responsibility gap: Ascribing responsibility for the actions of learning automata', *Ethics and Information Technology* 6: 175-183
- Mayer, M., 2015, 'The New Killer Drones: Understanding the Strategic Implications of Next-Generation Unmanned Combat Aerial Vehicles'. *International Affairs* 91, no.4: 765-780
- McAdams, R. H., 2015, *The Expressive Powers of Law: Theories and Limits*, Harvard University Press, Cambridge MA

- McArthur, N., 2017, 'The Case for Sexbots', in J. Danaher and N. McArthur (eds.) *Robot Sex: Social and Ethical Implications*, MIT Press, Cambridge
- McBride, M., (ed.), 2017, *New Essays on the Nature of Rights*, Oxford/Portland, OR: Hart Publishing
- McCarthy, L., 2004, 'Vicarious Liability in the Agency Context', *QUT Law Review*, [S.1.], v. 4, n. 2, dec. 2004. Available online: <https://lr.law.qut.edu.au/article/view/204> (accessed: 20th January 2026).
- McFarland, T. and McCormack, T., 2014, 'Mind the Gap: Can Developers of Autonomous Weapons Systems Be Liable for War Crimes?' *Int'l L. Stud.* 90:361
- McGlynn, C., 2017, 'Rape Trials and Sexual History Evidence: Reforming the Law on Third-Party Evidence', *The Journal of Criminal Law* 81(5): 367–392
- McKenna, M., 2005, 'Reasons Reactivity and Incompatibilist Intuitions', *Philosophical Explorations* 8(2): 131-143
- McKernan, B. and Davies, H., 2024, "The machine did it coldly": Israel used AI to identify 37,000 Hamas targets', *The Guardian*, Wed 3 Apr 2024 14.53 BST: <https://www.theguardian.com/world/2024/apr/03/israel-gaza-ai-database-hamas-airstrikes> (accessed: 19th January 2026).
- McKinnon, R., 2013, 'The supportive reasons norm of assertion', *American Philosophical Quarterly* 50(2): 121–135
- , 2014, 'Stereotype threat and attributional ambiguity for trans women', *Hypatia* 29(4): 857–872

——, 2015, *The Norms of Assertion: Truth, Lies, and Warrant*, Palgrave Macmillan: Houndmills

——, forthcoming, ‘How to be an optimist about aesthetic testimony’, *Episteme*

McMahan, J., 2009, *Killing in War* (OUP).

Medina, José, 2012, ‘Hermeneutical Injustice and Polyphonic Contextualisms: Social Silences and Shared Hermeneutical Responsibilities,’ *Social Epistemology: A Journal of Knowledge, Culture and Policy*, 26(2): 201–220

——, 2013, *The Epistemology of Resistance: Gender and Racial Oppression, Epistemic Injustice, and the Social Imagination*, Studies in Feminist Philosophy, New York

——, 2016. ‘Ignorance and Racial Insensitivity,’ *The Epistemic Dimensions of Ignorance*, R. Peels and M. Blaauw (ed.), Cambridge: Cambridge University Press, 178–201

Meier, W. M., 2019, ‘Lethal Autonomous Weapons Systems: Is It the End of the World as We Know It... Or Will We Be Just Fine?’, in C.M. Ford & W. S. Williams (eds.) *Complex Battlespaces*, Oxford University Press

Mele, A., 2011, ‘Moral Responsibility for Actions: Epistemic and Freedom Conditions’, *Philosophical Explorations* 13: 101-111

Merat, A., 2019, “The Saudis couldn’t do it without us’: the UK’s true role in Yemen’s deadly war’, *The Guardian*. Available online: <https://www.theguardian.com/world/2019/jun/18/the-saudis-couldnt-do-it-without-us-the-uks-true-role-in-yemens-deadly-war> (Accessed: 20th January 2026).

Mettraux, G., 2009, ‘The Law of Command Responsibility’, Oxford: Oxford University Press

Meyer, L. H., & Roser, D., 2010, 'Climate justice and historical emissions', *Critical Review of International Social and Political Philosophy*, 13(1):229–253. Available online: <https://doi.org/10.1080/13698230903326349> (accessed: 20th January 2026).

Meyer, S. S., 1993, 'Aristotle on Moral Responsibility: Character and Cause', *Issues in Ancient Philosophy*, Oxford: Blackwell

Mikkola, M., 2019, *Pornography: A Philosophical Introduction*, New York

Mill, J. S., 2011[2019], *On Liberty*, available online: <https://www.gutenberg.org/cache/epub/34901/pg34901-images.html> (accessed: 19th January 2026).

Miller, D., 2002, 'Distributing Responsibilities', *Journal of Political Philosophy*, 9:453-471. Available online: <https://doi.org/10.1111/1467-9760.00136> (accessed: 20th January 2026).

—, 2004, 'Holding Nations Responsible', *Ethics* 114(2): 240-268

—, 2007, 'Chapter 4 Two Concepts of Responsibility', *National Responsibility and Global Justice*, Oxford: OUP. Available online: <https://doi-org.ezproxy-prd.bodleian.ox.ac.uk/10.1093/acprof:oso/9780199235056.003.0004> (accessed: 21st January 2026).

Miller, S., 2016, 'Autonomous Weapons and Moral Responsibility', in S. Miller *Shooting to Kill: The Ethics of Police and Military Use of Lethal Force*

Mills, C., 2015, 'Race and Global Justice', in B. Buckinx, J. Trejo-Mathys, and T. Waligore,(eds.), *Domination and Global Political Justice*: 181-206, London: Routledge.

—, 2017, *Black Rights/White Wrongs: The Critique of Racial Liberalism*, Oxford: Oxford University Press

Ministry of Defence (MoD) 2011, *The UK Approach to Unmanned Aircraft Systems*, Wiltshire UK: Development, Concepts and Doctrine Centre

—, 2017, *Unmanned Aircraft Systems*, Joint Doctrine publication 0-30.2, Wiltshire UK: Development, Concepts and Doctrine Centre

Moore, M., 2000, 'The Metaphysics of Causal Intervention' *California Law Review* 88(3):827;

—, 2007, 'Causing, Aiding, and the Superfluity of Accomplice Liability' in *University of Pennsylvania Law Review* 156: 395-452

—, 2019, 'Causation in the Law', Zalta, E. N. (ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2019 Edition)

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L., 2016, 'The ethics of algorithms: Mapping the debate', *Big Data & Society*, 3(2), 1–21.

Morgan, P., 2012, 'Recasting Vicarious Liability', *The Cambridge Law Journal*, November 2012, Vol. 71, No. 3 (November 2012):615-650

Mori, M., 1970, 'The Uncanny Valley', *Energy*, Vol. 7, No. 4: 33–35 (in Japanese), K. MacDorman and N. Kageki (trans.), 2012, *IEEE ROBOTICS & AUTOMATION MAGAZINE*: 98-100

Morin, R., 2016, 'Can Child Dolls Keep Pedophiles from Offending?', *The Atlantic*, <https://www.theatlantic.com/health/archive/2016/01/can-child-dolls-keep-pedophiles-from-offending/423324/> (accessed: 19th January 2026).

Morrigan, V., 2023, 'Patriarchal Imaginaries Beyond the Human: 'Sex' Robots, Fetish and Fantasy in the Domination and Control of Women', in K. Richardson and C. Odland (eds.) *Man-Made Women*, Palgrave Macmillan

Morris, H., 1981, 'A Paternalistic Theory of Punishment' 18 *American Philosophical Quarterly* 263.

Munch, L., Mainz, J. and Bjerring, J.C., 2023, 'The value of responsibility gaps in algorithmic decision-making', *Ethics and Information Technology* 25:21. Available online: <https://philpapers.org/archive/MUNTVO-2.pdf> (accessed: 19th January 2026).

Murray, L., 2017, 'NAE WAY A sex robot has been given a SCOTTISH accent to sound less robotic', *The Scottish Sun*, 29 October 2017

Available online: <https://www.thescottishsun.co.uk/news/scottish-news/1766829/sex-robot-scottish-accent-sounds-less-robotic/> (accessed: 19th January 2026).

Naffine, N., 2009, *Law's Meaning of Life: Philosophy, Religion, Darwin and the Legal Person*, Hart Publishing

Nagel, T., 2012, *Mortal Questions*, Cambridge: Cambridge University Press

Nelkin, D.K. and Rickless, S.C., 2015, 'Review of Randolph Clarke's *Omissions: Agency, Metaphysics, and Responsibility*', *Notre Dame Philosophical Reviews* (February), 2015.02.33

—, 2017, 'Moral Responsibility for Unwitting Omissions: A New Tracing View', in D. Nelkin and S. Rickless (eds.), *The Ethics and Law of Omissions*, Oxford: Oxford University Press: 106–130

Neufeld, B.E. and C. Van Schoelandt, 2014, 'Political Liberalism, Ethos Justice, and Gender Equality,' *Law and Philosophy*, 33(1): 75–104

Neuner, J., 2013, 'Zur Rechtsfähigkeit des Anencephalus', 31 *Medizinrecht*

Nevejans, N., 2016, 'European Civil Law Rules in Robotics', Directorate General for Internal Policies 2016, Study for the JURI Committee PE 571.379

Nielly, C., 2020, 'Can We Let Algorithm Take Decisions We Cannot Explain? Towards Data', *Science*.

Nilsson, N., 2015, 'Introduction to Machine Learning: An Early Draft of a Proposed Textbook'. Available online: <https://ai.stanford.edu/~nilsson/MLBOOK.pdf> (accessed: 19th January 2026).

Nolan, D. and Davies, J., 2017, 'Torts and Equitable wrongs', in *English Private Law*, A. Burrows (ed.) (3rd edn. Oxford: Oxford University Press).

Noonan, J., 1976, *Persons and Masks of the Law*, New York: Farrar, Straus and Giroux

Norton Rose Fulbright, 2024, 'Artificial intelligence and liability: Key takeaways from recent EU legislative initiatives' (July 2024). Available online: <https://www.nortonrosefulbright.com/en/knowledge/publications/7052eff6/artificial-intelligence-and-liability> (accessed: 20th January 2026).

NSPCC 2023, 'Protecting children from sexual abuse', <https://learning.nspcc.org.uk/child-abuse-and-neglect/child-sexual-abuse> (accessed: 19th January 2026).

Nussbaum, M., 1995, 'Objectification', *Philosophy & Public Affairs* 24

——, 2011, *Creating Capabilities*, Cambridge, MA: Harvard University Press

Nyholm, S., 2018, 'Attributing agency to automated systems: Reflections on human-robot collaborations and responsibility-loci', *Science and Engineering Ethics*, 24(4):1201–1219

——, 2020, *Humans and Robots: Ethics, Agency, and Anthropomorphism*, Lanham: Rowman & Littlefield.

Odland, C. and Richardson, K., 2023, 'The End of Sex Robots—For the Dignity of Women and Girls', in K. Richardson and C. Odland (eds.) *Man-Made Women: The Sexual Politics of Sex Dolls and Sex Robots*, Palgrave Macmillan

Oosterheld, C., 2016, 'Backup utility functions as a fail-safe AI technique', in *Foundational Research Institute, Report FRI-16-2*

Okin, S., 1994, 'Political Liberalism, Justice, and Gender', *Ethics* 105(1): 23-43

——, 2004, 'Justice and Gender: An Unfinished Debate', *Fordham Law Review* 72: 1537-1567

Olivecrona, K., 1928, *Studier över begreppet juridisk person i romersk och modern rätt*, Appelbergs

Omar, R., 2017, 'Court of Appeal clarifies 'reasonable foreseeability test'', *Forbes Solicitors*. Available online: <https://www.forbessolicitors.co.uk/articles/court-of-appeal-clarifies-reasonable-foreseeability-test> (accessed: 8th January 2026).

ONS, 2021, 'Sexual offences in England and Wales overview: year ending March 2020', Available online: <https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/bulletins/sexualoffencesinenglandandwalesoverview/march2020> (accessed: 19th January 2026).

Open Letter to the EU Commission, 2018, 'OPEN LETTER TO THE EUROPEAN COMMISSION: ARTIFICIAL INTELLIGENCE AND ROBOTICS', Available online: <http://www.robotics-openletter.eu/> (accessed: 19th January 2026).

Orts, E., 2013, *Business Persons. A Legal Theory of the Firm*, OUP

Otsuka, M., 1998, 'Incompatibilism and the Avoidability of Blame', *Ethics* 108: 685-701

Page, E., 2011, 'Climatic Justice and the Fair Distribution of Atmospheric Burdens', *The Monist* 94(3):412-432.

——, 2012, 'Give it up for Climate Change: A Defence of the Beneficiary Pays Principle', *International Theory* 4:300-330.

- Pallikkathayil, J., 2016, 'Revisiting the Interest Theory of Rights: Discussion of The Morality of Freedom', *Jerusalem Review of Legal Studies*, 14(1): 147–157
- Pangotra, A., 2025, 'UK to Criminalise AI-Generated Sexual Abuse Images Soon', *CyberPeace*. Available online: <https://cyberpeace.org/resources/blogs/uk-to-criminalise-ai-generated-sexual-abuse-images-soon> (accessed: 19th January 2026).
- Parikh, R., 2002, 'Social Software', *Synthese* 13(2): 187-211
- Patridge, S., 2010, 'The Incorrigible Social Meaning of Video Game Imagery', *Ethics and Information Technology* 13(4): 303-312
- , 2013, 'Pornography, Ethics and Video Games', *Ethics and Information Technology* 15(1)L 25-34
- Patridge, S. and Jordan, A., 2010, 'Against the moralistic fallacy: In Defence of Modest Moralism about Humour', *Ethical Theory and Moral Practice* 15(1): 84-95
- Peeters, A. and Haselager, P., 2021, 'Designing Virtuous Sex Robots', *International Journal of Social Robotics* 13: 55-66
- Pesch, U., 2015, 'Engineers and Active Responsibility', *Science and Engineering Ethics*, 21(4), 925–939.
- Plunkett, J., 2016, 'Taking Stock of Vicarious Liability' 132 *LQR* 556.
- Priest, G., 1985, 'The Invention of Enterprise Liability: A Critical History of the Intellectual Foundations of Modern Tort Law', *J. Legal Stud.* 14:461.
- , 1987, 'The Current Insurance Crisis and Modern Tort Law', *Yale L.J.* 96 :1521
- Purves, D., Jenkins, R. and Strawser, B.J., 2015, 'Autonomous machines, moral judgment, and acting for the right reasons', *Ethical Theory and Moral Practice* 18: 851-872

Queensland Law Reform Commission, 2001, *Vicarious Liability*, Report No 56.

Quinn, W., 1993, 'The Right to Threaten and the Right to Punish' in *Morality and Action* (Cambridge: CUP, 1993)..

Quong, J., 2004, 'The Scope of Public Reason', *Political Studies* 52: 233-250

——, 2011, *Liberalism Without Perfection*, Oxford: Oxford University Press.

——, 2012, 'Liberalism Without Perfection: Replies to Gaus, Colburn, Chan, Bocchiola,' *Philosophy & Public Issues*, 2(1): 51–79.

——, 2013, 'On The Idea of Public Reason,' in *The Blackwell Companion to Rawls*, J. Mandle and D. Reidy (eds.), Oxford: Wiley-Blackwell: 265–280.

——, 2014, 'What is the Point of Public Reason?' *Philosophical Studies*, 170(3): 545–553.

Racco, M., 2018, 'Will having a sex robot in the home pose a risk to children?', *Global News* (17 July 2018). Available online: <https://globalnews.ca/news/4336151/sex-robot-kids-risks/> (accessed: 4th January 2026).

Radomirescu, A., 2022, 'Attributing Responsibility for Autonomous Weapons', Thesis. Personal communication.

Rage Against The Machine, 1992, 'Settle for Nothing', *Rage Against The Machine*. Available online: <https://genius.com/Rage-against-the-machine-settle-for-nothing-lyrics> (accessed: 22nd January 2026).

Rahwan, I., 2018, 'Society-in-the-loop: Programming the algorithmic social contract', *Ethics and Information Technology*, 20(1):5–14.

Rawls, J., 1993[2005], *Political Liberalism*, Columbia Classics in Philosophy

——, 1995, 'Reply to Habermas,' *The Journal of Philosophy*, 92(3): 132–180

——, 1997, ‘The Idea of Public Reason Revisited’, *The University of Chicago Law Review*, Vol. 64, No. 3: 765-807

——, 1999a, *A Theory of Justice: Revised Edition*, Oxford: Oxford University Press

——, 1999b, *Collected Papers*, S. Freeman (ed.), Cambridge, MA: Harvard University Press

——, 2001, *Justice as Fairness: A Restatement*, Cambridge, MA: Harvard University Press

Raytheon ‘Phalanx’, *Phalanx Weapon System*. Available online: <https://www.rtx.com/raytheon/what-we-do/sea/phalanx-close-in-weapon-system> (accessed: 5th January 2026).

Raz, Joseph, 1970, *The Concept of a Legal System: An Introduction to the Theory of Legal System*, Oxford: Clarendon Press

——, 1986, *The Morality of Freedom*, Oxford: Clarendon Press

——, 1994, *Ethics in the Public Domain: Essays in the Morality of Law and Politics*, Oxford: Clarendon Press; reprinted, 2001

——, 1995, *Legal Rights*, in ETHICS IN THE PUBLIC DOMAIN: ESSAYS IN THE MORALITY OF LAW AND POLITICS 259

Realbotix, 2023, <https://realbotix.com/> (accessed: 19th January 2026)

Report from the Expert Group on Liability and New technologies, 2019, ‘Liability for Artificial Intelligence and other emerging technologies’. Available online: https://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/COMMITTEES/JURI/DV/2020/01-09/AI-report_EN.pdf (accessed: 19th January 2026).

Restatement (Third) of Agency, 2006, Available online: <https://studylib.net/doc/8168533/restatement-of-the-law---agency-restatement--third--of-a> (accessed: 19th January 2026).

Ribeiro, M., Singh, S. and Guestrin, C., 2016, 'Why Should I Trust You?' *Explaining the Predictions of Any Classifier* <https://arxiv.org/pdf/1602.04938.pdf> (accessed: 19th January 2026).

Richardson, K., 2015, 'The Asymmetrical 'Relationship': Parallels Between Prostitution and the Development of Sex Robots', *SIGCAS Computers & Society* 45(3): 290-293

——, 2016, 'Sex robot matters: slavery, the prostituted, and the rights of machines', *IEEE Technology and Society Magazine* 35(2): 46-53

——, 2023, 'The End of Sex Robots: Porn Robots and Representational Technologies of Women and Girls', in K. Richardson and C. Odland (eds.) *Man-Made Women: The Sexual Politics of Sex Dolls and Sex Robots*, Palgrave Macmillan

Richardson, K., and C. Odland (eds.), 2023, *Man-Made Women: The Sexual Politics of Sex Dolls and Sex Robots*, Palgrave Macmillan

Riehm, T., 2020, 'Riehm: Nein zur Person!', *Recht Digital* 42

Rissman, K., 2025, 'Judge rules possession of AI-generated child porn could be protected under First Amendment', *Independent*. Available online: <https://www.independent.co.uk/news/world/americas/crime/child-abuse-ai-judge-charges-b2717291.html> (accessed: 24th January 2026).

Roberts, J., 1989, 'Aristotle on Responsibility for Action and Character', *Ancient Philosophy*

Robeyns, I. and Byskov, M., 'The Capability Approach', *The Stanford Encyclopedia of Philosophy* (Summer 2023 Edition), Edward N. Zalta & Uri Nodelman (eds.)

Robichaud, P. and Wieland, J.W., 2017, *Responsibility: The Epistemic Condition*, Oxford.

Robillard, M., 2018, 'No Such Thing as Killer Robots', *Journal of Applied Philosophy* Vol. 35, No. 4

Robinson, P.H., 2013, *Intuitions of Justice and the Utility of Desert*, Oxford: OUP.

Robinson, P.H. and Kurzban, R., 2007, 'Concordance and Conflict in Intuitions of Justice', *Minnesota Law Review* 91: 1829.

Robinson, P.H. Kurzban, R. and Jones, O.D., 2007, 'The Origins of Shared Intuitions of Justice', *Vanderbilt Law Review* 60: 1633.

Rochfeld, J., 2013, *Les grandes notions du droit privé*, 2nd ed. Presses Universitaires de France

Roff, H., 2013, 'Killing in War: Responsibility, Liability and Lethal Autonomous Robots' in F. Allhoff, N. G. Evans and A. Henschke (eds.) *Routledge Handbook of Ethics and War: Just War Theory in the Twenty-First Century*, New York: Routledge 352-364

Roff, H. and Moyes, R., 2016, Article 36, 'Meaningful Human Control, Artificial Intelligence and Autonomous Weapons', UN GGE LAWs briefing paper

Rome Statute of the International Criminal Court 1998, art. 75(2), July 17, 1998, 2187 U.N.T.S. 90

Rudy-Hiller, F., 2025, 'The Epistemic Condition for Moral Responsibility', *The Stanford Encyclopedia of Philosophy* (Fall 2025 Edition), Edward N. Zalta & Uri Nodelman (eds.), Available online: <https://plato.stanford.edu/archives/fall2025/entries/moral-responsibility-epistemic/> (accessed: 20th January 2026).

Russell, S., 2021, 'It's time to ban autonomous killer robots before they become a threat', Financial Times August 5 2021, <https://www.ft.com/content/04a07148-d963-4886-83f6-fcaf4889172f>

Russell, S. J., and Norvig, P., 2009, *Artificial Intelligence: A Modern Approach*, 3rd ed. Prentice Hall Press

SAE, 2021, 'Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles', Society for Automotive Engineers, Available online: https://www.sae.org/standards/content/j3016_202104/ (accessed: 19th January 2026).

Saerbeck, M., and Bartneck, C., 2010, 'Attribution of Affect to Robot Motion', PROCEEDINGS OF THE 5TH ACM/IEEE INTERNATIONAL CONFERENCE ON HUMAN-ROBOT INTERACTION (HRI2010), Osaka: 53-60

Salmond, J. W., 1920, *Jurisprudence*, Sweet and Maxwell

Sandberg, A., 2015, 'Sex and death among the robots: When should we campaign to ban robots?', *Practical Ethics Blog*

Sandel, M., 2012, *What Money Can't Buy*, London: Penguin

Santoni de Sio, F., & Mecacci, G., 2021, 'Four Responsibility Gaps with Artificial Intelligence: Why they Matter and How to Address them', *Philosophy & Technology*, 34: 1057-1084

Santoni de Sio, F., & van den Hoven, J., 2018, 'Meaningful Human Control over Autonomous Systems: A Philosophical Account', *Front Robot AI*:5:15.

Sartorio, C., 2016, *Causation and Free Will*, Oxford: Oxford University Press

——, 2017, ‘Ignorance, Alternative Possibilities, and the Epistemic Conditions for Responsibility’, in R. Peels (ed.), *Perspectives on Ignorance from Moral and Social Philosophy*, New York: Routledge.: 15–29.

Savigny, F. C. von, 1840, *System des heutigen Römischen Rechts*, Veit

——, 1867, *System of the Modern Roman law, Volume 1*, W. Holloway (tr.), J. Higginbotham (ed.)

——, 1884, *Jural Relations; or, the Roman Law of Persons as Subjects of Jural Relations: Being a Translation of the Second Book of Savigny’s System of Modern Roman Law*, W. H. Rattigan (tr., ed.), Wildy & Sons

Scanlon, T. M., 1998, *What We Owe to Each Other*, Harvard University Press

——, 1998[2000], *What We Owe to Each Other*. Cambridge, MA and London: Belknap Press of Harvard University Press.

Schellekens, M., 2018, ‘No-fault compensation schemes for self-driving vehicles’, *Law, Innovation and Technology*, 10(2), 314–333. Available online: <https://doi.org/10.1080/17579961.2018.1527477> (accessed: 19th January 2026).

Scherer, M., 2017, ‘Is AI Personhood Already Possible Under U.S. LLC Laws? (Part One: New York)’, 14 May 2017

Schmitt, F., 2006, ‘Testimonial justification and transindividual reasons’, *The Epistemology of Testimony*, J. Lackey and E. Sosa (eds.) Oxford: Oxford University Press: 193–224

Schouten, G., 2013, ‘Restricting Justice: Political Interventions in the Home and in the Market,’ *Philosophy & Public Affairs*, 41(4): 357–388.

——, 2017, ‘Citizenship, Reciprocity, and the Gendered Division of Labor: A Stability Argument for Gender Egalitarian Interventions,’ *Politics, Philosophy, & Economics*, 16(2): 174–209.

——, 2019, *Liberalism, Neutrality, and the Gendered Division of Labor*, New York: Oxford University Press

Schwarz, E., 2025, ‘Engineering Moral Failure? The Challenges of Algorithmic Ethics for Lethal Autonomous Weapon Systems’, in *The Realities of Autonomous Weapons*: 232–258.

Scott, H., 2019, ‘The History of Foreseeability’, *Current Legal Problems*, Vol.72(1):287–314. Available online: <https://doi.org/10.1093/clp/cuz009> (accessed: 20th January 2026).

Second Restatement of Torts, 1965, *American Law Institute*, St. Paul, Minn.: American Law Institute Publishers

Seixas-Nuñez, SJ, A., 2022a, ‘Autonomous weapons systems and the procedural accountability gap’, *Brooklyn Journal of International Law*, 46(2): 421–478. Available online: <https://brooklynworks.brooklaw.edu/cgi/viewcontent.cgi?article=1973&context=bjil> (accessed: 19th January 2026).

——, 2022b, *The legality and accountability for autonomous weapon systems. A humanitarian law perspective*, Cambridge University Press.

——, 2024, ‘Scapegoats! Assessing the Liability of Programmers and Designers for Autonomous Weapons Systems’, in J.M. Schraagen (ed.) *Responsible Use of AI in Military Systems*, CRC Press LLC. Available online: <http://ebookcentral.proquest.com/lib/oxford/detail.action?docID=31281995> (accessed: 19th January 2026).

Sen, A., 1992, 1992, *Inequality Re-examined*, Oxford: Clarendon Press

Sexual Offences Act 2003, Available online: <https://www.legislation.gov.uk/ukpga/2003/42/part/1/crossheading/rape> (accessed: 19th January 2026).

Shabo, S., 2015, 'More Trouble with Tracing', *Erkenntnis*, 80(5): 987–1011.

Sharkey, C., 2024, 'Products Liability for Artificial Intelligence', *LawFare* (Sep. 25, 2024). Available online: <https://www.lawfaremedia.org/article/products-liability-for-artificial-intelligence> (accessed: 20th January 2026).

Sharkey, N., 2010, 'Saying 'no!' to lethal autonomous targeting', *Journal of Military Ethics*, 9(4), 369–383.

—, 2012, 'The evitability of autonomous robot warfare', *International Review of the Red Cross* Vol.94, No.886 Summer 2012. Available online: <https://international-review.icrc.org/sites/default/files/irrc-886-sharkey.pdf> (accessed: 19th January 2026).

Sharkey, N., van Wynsberghe, A., Robbins, S., and Hancock, E., 2017, 'Our sexual future with robots', The Hague: Foundation for Responsible Robotics. Available online: <https://www.scribd.com/document/353020582/Report-Our-Sexual-Future-With-Robots> (accessed: 19th January 2026).

Shen, F. X., 2019, 'Government officials have no idea how to regulate the growing sex-robot industry', *The Business Insider*, Feb. 12 2019

Available online: <https://www.businessinsider.com/sex-robot-industry-regulations-2019-2?r=US&IR=T> (accessed: 19th January 2026).

Sheiner, N., 1977, 'DSS and a Proposed Theory of Enterprise Liability' (1977–8), 46 *Fordham Law Review* 963.

Shoemaker, D., 2011, 'Attributability, Answerability, and Accountability: Toward a Wider Theory

of Moral Responsibility,' *Ethics*, 121(3), 602–632. DOI: 10.1086/659003.

———, 2013, 'On Criminal and Moral Responsibility,' in M. Timmons (ed.), *Oxford Studies in*

Normative Ethics: Vol.3:154–178, Oxford: Oxford University Press.

———, 2015, *Responsibility from the Margins*, Oxford: Oxford University Press.

Shor, E. and Seida, K., 2019, "Harder and Harder?' Is Mainstream Pornography Becoming Increasingly Violent and Do Viewers Prefer Violent Content?', *The Journal of Sex Research* 56: 16–28

———, 2020, *Aggression in Pornography: Myths and Realities*, Routledge

Sieg, R., 2005, 'Attempted Possession of Child Pornography-A Proposed Approach for Criminalizing Possession of Child Pornographic Images of Unknown Origin', 36 U. TOL. L. Rev. 263.

Sierra Leone Statute, Statute of the Special Court for Sierra Leone art. 6(3), Jan. 16, 2002, 2178 U.N.T.S. 145

Simester, A. *et al.*, 2016, *Simester and Sullivan's Criminal Law: Theory and Doctrine* (London: Bloomsbury, 2016)

Simmonds, N. E., 1998, 'Rights at the Cutting Edge', in Kramer, Simmonds, and Steiner 1998: 113–232

Simpson, T. and Müller, V., 2016, 'Just War and Robots' Killings', *The Philosophical Quarterly* 66(263):306.

- Singer, P., 2009, 'Wired for War: The Robotics Revolution and Conflict in the 21st Century', New York: Penguin
- Sinnott-Armstrong, W. and Skorburg, J. A., 2021, 'How AI can aid bioethics', *Journal of Practical Ethics*
- Sliwa, P., 2019, 'The Power of Excuses,' *Philosophy and Public Affairs* 47: 37–71.
- Smith, A., 2005, 'Responsibility for Attitudes: Activity and Passivity in Mental Life,' *Ethics*, 115(2):236–271.
- , 2012, 'Attributability, Answerability, and Accountability: In Defense of a Unified Account,' *Ethics*, 122(3):575–589.
- , 2015, 'Responsibility as Answerability,' *Inquiry*, 58(2):99–126.
- Smith, B., 1928, 'Legal Personality', *The Yale Law Journal* Vol. 37, No. 3: 283-299
- Smith, G., et al., 2024, 'Liability for Harms from AI Systems', Rand (Nov. 24, 2024). Available online: https://www.rand.org/pubs/research_reports/RRA3243-4.html (accessed: 20th January 2026).
- Smith, O. and Skinner, T., 2017, 'How rape myths are used and challenged in rape and sexual assault trials', *Social & Legal Studies*
- Smith, S., 2003, 'Rational Capacities, or: How to Distinguish Recklessness, Weakness, and Compulsion', in Stroud, S. and Tappolet, C. (eds.) *Weakness of Will and Practical Irrationality*: 17-38, New York NY: Oxford University Press
- Solovyeva, A. & Hynek, N., 2023, 'When stigmatization does not work: over-securitization in efforts of the Campaign to Stop Killer Robots', *AI and Society* 38 (6):2547-2569.
- Solum, L. B., 1992, 'Legal Personhood for Artificial Intelligences', 70 NC L Rev 1231

Spadaro, A., 2023, 'A Weapon is no Subordinate. Autonomous Weapon Systems and the Scope of

Superior Responsibility', in *Journal of International Criminal Justice*: 1-18.

Sparrow, R., 2004, 'The Turing Triage Test', *Machine Ethics and Robot Ethics*: 95-105

——, 2007, 'Killer Robots', *Journal of Applied Philosophy*, 24: 62-77

——, 2016, 'Robots and Respect: assessing the case against autonomous weapon systems', *Ethics & International Affairs* 30: 93-116

——, 2017, 'Robots, rape, and representation', *International Journal of Social Robotics* 9(4): 465-477

——, 2019, 'Regulating Child Sex Robots: Restriction or Experimentation?', *Med Law Rev* 27(4): 553-575

Srinivasan, A., 2021, *The Right to Sex: Feminism in the Twenty-First Century*, New York: Farrar, Straus and Giroux

Stapleton, J., 1986, 'Products Liability Reform--Real or Illusory?', *Oxford Journal of Legal Studies*, Winter, 1986, Vol. 6, No. 3 (Winter, 1986): 392-422

Starr, E., 1970, 'War'. Available online: <https://genius.com/Edwin-starr-war-lyrics> (accessed: 21st January 2026).

Steiner, H., 1994, *An Essay on Rights*, Oxford/Cambridge, MA: Blackwell

Stemplowska, Z., 2020, 'Duties to the Dead: Is Posthumous Mitigation of Injustice Possible?', *Oxford Studies in Political Philosophy* 6, Oxford University Press

Sterri, A. and Earp, B., 2021, 'The ethics of sex robots', in *The Oxford Handbook of Digital Ethics*, Oxford: Oxford University Press

- Stevens, R., 2007, *Torts and Rights*, Oxford: Oxford University Press
- Steward, H., 2012, *A Metaphysics for Freedom* (Oxford, 2012; online edn, Oxford Academic, 24 May 2012). Available online: <https://doi.org/10.1093/acprof:oso/9780199552054.001.0001>, (accessed: 20th Jan. 2026).
- Stilgoe, J., 2020, 'Who Killed Elaine Herzberg?' In *Who's Driving Innovation?* (pp. 1–6). Springer International Publishing. Available online: https://doi.org/10.1007/978-3-030-32320-2_1 (accessed: 19th January 2026).
- Stone, C., 1972[2010], 'Should Trees Have Standing?', in *Should Trees Have Standing? Law, Morality, and the Environment*, 3rd edition, OUP
- Strawson, P., 1962[2008], 'Freedom and Resentment', in *Freedom and Resentment and Other Essays*, London: Routledge.
- Strikwerda, L., 2017, 'Legal and Moral Implications of Child Sex Robots', in J. Danaher and N. McArthur (eds.) *Robot Sex: Social and Ethical Implications*, MIT Press, Cambridge
- Sturges, F., 2020, 'Sex Robots & Vegan Meat by Jenny Kleeman review – the future of food, birth and death?', *The Guardian*, 1 July 2020. Available online: <https://www.theguardian.com/books/2020/jul/01/sex-robots-vegan-meat-by-jenny-kleeman-review-the-future-of-food-birth-and-death> (accessed: 19th January 2026).
- Sullivan, G. R., 1996, 'Making Excuses' in A. P. Simester and A. T. H. Smith *Harm and Culpability* (Oxford: OUP, 1996).
- Sumner, L., 1987, *The Moral Foundations of Rights*, Oxford: Clarendon Press.
- Sun, C., Bridges, A., Johnson, J. A. and Ezzell, M. B., 2016, 'Pornography and the Male Sexual Script: An Analysis of Consumption and Sexual Relations', *Archives of Sexual Behavior*, 45: 983–94

Sunstein, C., 1996, 'On the Expressive Function of Law', *University of Pennsylvania Law Review*, Vol. 144, No. 5: 2021-2053

Sutton, R. S., Barto, A. G., 1998, *Reinforcement Learning: An Introduction*, Cambridge: MIT Press.

Taddeo, M. and Blanchard, A., 2021, 'A Comparative Analysis of the Definitions of Autonomous Weapons Systems'. Available online: <https://documents.unoda.org/wp-content/uploads/2021/10/20210721-Autonomous-Weapon-Systems-Definitions-TO-SHARE.pdf> (accessed: 5th January 2026).

Tadros, V., 2011, 'Against Desert', *The Ends of Harm: The Moral Foundations of Criminal Law*, Oxford Legal Philosophy (Oxford, 2011; online edn, Oxford Academic, 19 Jan. 2012). Available online: <https://doi-org.ezproxy-prd.bodleian.ox.ac.uk/10.1093/acprof:oso/9780199554423.003.0004> (accessed: 5th January 2026).

———, 2016, *Wrongs and Crimes*, OUP

Talbert, M., 2024, 'Moral Responsibility', *The Stanford Encyclopedia of Philosophy* (Fall 2025 Edition), E. N. Zalta & U. Nodelman (eds.). Available online: <https://plato.stanford.edu/archives/fall2025/entries/moral-responsibility/> (accessed: 20th January 2026).

Tan, J., 2022, 'Vicarious liability as risk-through-placing', *Common Law World Review*, 51(4):268-295. Available online: <https://doi.org/10.1177/14737795221116397> (accessed: 20th January 2026)

Tan, K.-C., 2007, 'Colonialism, Reparations and Global Justice' in Miller, J., and Kumar, R., (eds.), *Reparations: Interdisciplinary Inquiries*: 280-306, Oxford: Oxford University Press

Tan, S. H., 2002, 'Validity and Obligation in Natural Law Theory', *Regent University Law Review* 15: 195-221

Taylor Wessing, 2025, 'A new era for product liability in the EU', *Synapse* July 2025. Available online: <https://www.taylorwessing.com/en/synapse/2025/post-market-activities-including-pms-and-product-liability/a-new-era-for-product-liability-in-the-eu> (accessed: 20th January 2026).

The Law Commission and the Scottish Law Commission, 1977, 'Liability for defective products: final report', 1 June 1977. Available online: <https://lawcom.gov.uk/publication/liability-for-defective-products-final-report/> (accessed: 20th January 2026).

Tigard D.W., 2020, 'There is no techno-responsibility gap', *Philosophy & Technology* Vol.34:589-607. Available online: <https://doi.org/10.1007/s13347-020-00414-7> (accessed: 19th January 2026).

Timmins, B., 2017, 'New sex robots with 'Frigid' settings allow men to simulate rape', *The Independent* 19 July, <https://www.independent.co.uk/life-style/sex-robots-frigid-settings-rape-simulation-men-sexual-assault-a7847296.html> (accessed: 4th January 2026)

Tollefsen, D., 2002, 'Collective intentionality and the social sciences', *Philosophy of the Social Sciences* 32(1): 25-50

——, 2015, *Groups as agents*, Polity Press

Tomlin, P., 2017, 'Innocence Lost: A Problem for Punishment as Duty', *Law and Philosophy* 36(3):225-254. Available online: <https://wrap.warwick.ac.uk/id/eprint/104424/1/WRAP-innocence-lost-Tomlin-2017.pdf> (accessed: 20th January 2026).

Tonkens, R., 2012, 'Out of character: on the creation of virtuous machines', *Ethics of Information Technology* 14(2): 137-149

Tort Law of the People's Republic of China, 2009. Available online: https://english.mofcom.gov.cn/Policies/LawsRegulations/art/2013/art_e796d185d3f243a7a292ee94c0e29fc7.html (accessed: 14th January 2026).

Trabucco, L., 2025, 'AI-Enabled Autonomous Weapons and Human Control: Part I: Human Control and Machine Learning Design and Development', *International Law Studies* Vol.106: 534

Trudy, J., Brown, R., Ibrahim, I. and Parellada, O., 2021, 'A Sandbox Approach to Regulating High-Risk Artificial Intelligence Applications', *European Journal of Risk Regulation*

Turing, A., 1950, 'Computing machinery and intelligence', *Mind* 59(236): 433-460

Turkle, S., 2010, 'In Good Company?: On the Threshold of Robotic Companions', in *Close Engagements with Artificial Companions: Key Social, Psychological, Ethical and Design Issues*, Y. Wilks (ed.), John Benjamins Publishing Company

Turner, J., 2019, *Robot Rules: Regulating Artificial Intelligence*, Palgrave Macmillan

UK Parliament Home Affairs Committee, 2022, 'Investigation and prosecution of rape', *Eighth Report of Session 2021-22*, Available online: <https://committees.parliament.uk/publications/9600/documents/166175/default/> (accessed: 4th January 2026).

UN, 2019, 'Report of the 2019 session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems. CCW/GGE.1/2019/3, 25 September 2019, Annex IV.

—, 2023a, ‘First Committee Approves New Resolution on Lethal Autonomous Weapons, as Speaker Warns ‘An Algorithm Must Not Be in Full Control of Decisions Involving Killing’’, GA/DIS/3731, 1 November 2023. Available online: <https://press.un.org/en/2023/gadis3731.doc.htm> (accessed: 19th January 2026).

—, 2023b, ‘Resolution adopted by the General Assembly on 22 December 2023: 78/241. Lethal autonomous weapons systems’. Available online: <https://documents.un.org/doc/undoc/gen/n23/431/11/pdf/n2343111.pdf> (accessed: 19th January 2026).

—, 2024, ‘Current developments in science and technology and their potential impact on international security and disarmament efforts’, *Report of the Secretary-General*, 23 July 2024. Available online: <https://docs.un.org/en/A/79/224> (accessed: 19th January 2026).

UN GGE, 2023, ‘Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects’, *Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons System* (Geneva, 6-10 March, and 15-19 May 2023). Available online: [https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons_-_Group_of_Governmental_Experts_on_Lethal_Autonomous_Weapons_Systems_\(2023\)/CCW_GGE1_2023_CRP.1_0.pdf](https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons_-_Group_of_Governmental_Experts_on_Lethal_Autonomous_Weapons_Systems_(2023)/CCW_GGE1_2023_CRP.1_0.pdf) (accessed: 5th January 2026).

—, 2024, ‘GGE on LAWS Rolling text, status date: 8 November 2024’. Available online: [https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons_-_Group_of_Governmental_Experts_on_Lethal_Autonomous_Weapons_Systems_\(2024\)/Revised_rolling_text_as_of_8_November_2024_final.pdf](https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons_-_Group_of_Governmental_Experts_on_Lethal_Autonomous_Weapons_Systems_(2024)/Revised_rolling_text_as_of_8_November_2024_final.pdf) (accessed: 5th January 2026).

—, 2025, ‘GGE on LAWS Rolling text, status date: 12 May 2025’. Available online: https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons_-_

[Group of Governmental Experts on Lethal Autonomous Weapons Systems \(2025\)/Revised rolling text as of 12 May 2025.pdf](#) (accessed: 5th January 2026).

UN Libya Report 2021, *Letter dated 8 March 2021 from the Panel of Experts on Libya established pursuant to resolution 1973 (2011) addressed to the President of the Security Council*. Available online: <https://documents-dds-ny.un.org/doc/UNDOC/GEN/N21/037/72/PDF/N2103772.pdf?OpenElement> (accessed: 4th January 2026).

UN Secretary-General, 2025, ‘Conflict-related sexual violence: report of the United Nations Secretary-General’, S/2025/389 (15 July 2025). Available online: <https://www.un.org/sexualviolenceinconflict/wp-content/uploads/2025/08/report/report-of-the-secretary-general-on-conflict-related-sexual-violence/SG-Report-2024-FINAL.pdf> (accessed: 4th January 2026).

UN Timeline, 2023, ‘Available online: <https://archive-disarmament.unoda.org/timeline-of-laws-in-the-ccw/>’ (accessed: 5th January 2026).

UN Women, 2025, ‘Facts and figures: Ending violence against women’, (19 November 2025). Available online: <https://knowledge.unwomen.org/en/articles/facts-and-figures/facts-and-figures-ending-violence-against-women> (accessed: 4th January 2026).

UNICEF, 2025, ‘Sexual violence’. Available online: <https://data.unicef.org/topic/child-protection/violence/sexual-violence/> (accessed: 4th January 2026).

Universal Declaration of Human Rights. Available online: <https://www.un.org/en/about-us/universal-declaration-of-human-rights> (accessed: 20th January 2026).

Upton, J., Hazell, A., Abbott, R., and Pilling K., 2021, ‘Independent report: The relationship between pornography use and harmful sexual behaviours’. Available online:

<https://www.gov.uk/government/publications/research-publications/the-relationship-between-pornography-use-and-harmful-sexual-behaviours#the-relationship-between-pornography-use-and-harmful-sexual-behaviours> (accessed: 19th January 2026).

US Department of Defense DoD 2012, *Directive Number 3000.09 on Autonomy in Weapon Systems*, November 21 2012
<https://www.esd.whs.mil/portals/54/documents/dd/issuances/dodd/300009p.pdf> (accessed: 19th January 2026).

—, 2018, ‘Unmanned Systems Integrated Roadmap 2017-2042’. Available online: <https://apps.dtic.mil/sti/pdfs/AD1059546.pdf> (accessed: 19th January 2026).

—, 2023, ‘DOD DIRECTIVE 3000.09 AUTONOMY IN WEAPON SYSTEMS’, 25 January 2023, Office of the Under Secretary of Defense for Policy. Available online: <https://www.esd.whs.mil/portals/54/documents/dd/issuances/dodd/300009p.pdf> (accessed: 19th January 2026).

US Senate Committee on the Judiciary, 2025, ‘Durbin, Hawley Introduce Bill Allowing Victims To Sue AI Companies’, 29 September 2025. Available online: <https://www.judiciary.senate.gov/press/dem/releases/durbin-hawley-introduce-bill-allowing-victims-to-sue-ai-companies> (accessed: 20th January 2026).

Vadas, M., 2005, ‘The Manufacture-for-Use of Pornography and Women’s Inequality’, *Journal of Political Philosophy* 13: 174-193

Vallentyne, P., 2007, ‘Libertarianism and the state’, *Social Philosophy and Policy*, 24: 187–205

Vallier, K., 2011, ‘Against Public Reason Liberalism’s Accessibility Requirement,’ *Journal of Moral Philosophy*, 8(3): 366–389.

—, 2014, *Liberalism and Public Faith: Beyond Separation*, New York: Routledge.

—, 2016, ‘Public Justification vs Public Deliberation: The Case for Divorce,’ *Canadian Journal of Philosophy*, 45(2): 139–158.

—, 2017, ‘On Jonathan Quong’s Sectarian Political Liberalism,’ *Criminal Law and Philosophy*, 11(1): 175–194.

—, 2019, *Must Politics Be War? Restoring Our Trust in the Open Society*, New York: Oxford University Press

Vallor, S., 2016, ‘Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting’, Oxford University Press, Oxford

Van de Poel, I., 2011, ‘The Relation Between Forward-Looking and Backward-Looking Responsibility’, in: N. Vincent, I. van de Poel, J. van den Hoven, (eds) *Moral Responsibility, Library of Ethics and Applied Philosophy*, vol 27. Springer, Dordrecht. Available online: https://doi.org/10.1007/978-94-007-1878-4_3 (accessed: 20th January 2026).

Van de Poel, I., & Sand, M., 2018, ‘Varieties of responsibility: two problems of responsible innovation’, *Synthese*. Vol.198:4769-4787. Available online: <https://doi.org/10.1007/s11229-018-01951-7> (accessed: 19th January 2026).

van Inwagen, P., 1978, ‘Ability and Responsibility’, *Philosophical Review* 87: 201-224

—, 1983, *An Essay on Free Will*, New York NY: Oxford University Press

Vansweevelt, T., & Weyts, B., 2009, *Handboek buitencontractueel aansprakelijkheidsrecht*, Intersentia.

Vargas, M., 2005, ‘The trouble with tracing’, *Midwest Studies in Philosophy* 29(1): 269-291

Vargo, J., 2014, PRODUCTS LIABILITY PRACTICE GUIDE, *LEXIS ADVANCE*. Available online: <https://advance.lexis.com/api/toc/source/MTA2ODYyNQ?origination=lawcc> (accessed: 20th January 2026).

Veluwenkamp, H., 2025, 'What responsibility gaps are and what they should be', *Ethics and Information Technology* 27 (1):1-13. Available online: <https://philpapers.org/go.pl?id=VELWRG&proxyId=&u=https%3A%2F%2Fdx.doi.org%2F10.1007%2Fs10676-025-09823-8> (accessed: 19th January 2026).

Veprikov, A., Afanasyev, A., Khritankov, A., 2025, 'A mathematical model of the hidden feedback loop effect in machine learning systems', *Knowledge and Information Systems*, 67: 11425–11454.

Vera-Gray, F., McGlynn, C., Kureshi, I., and Butterby, K., 2021, 'Sexual violence as a sexual script in mainstream online pornography', *The British Journal of Criminology* 61: 1243-1260

Vienna Convention on the Law of Treaties, 1969, Available online: https://legal.un.org/ilc/texts/instruments/english/conventions/1_1_1969.pdf (accessed: 19th January 2026).

von Eschenbach, W., 2021, 'Transparency and the Black Box Problem: Why We Do Not Trust AI', *Philosophy and Technology* 34(4):1607–1622

WAFEW, 2021, 'Written evidence submitted by Women's Aid Federation of England' (INV0010), Available online: <https://committees.parliament.uk/writtenevidence/36735/pdf/> (accessed: 19th January 2026).

Wagner, G., 2016, 'Autonomous Weapon Systems', *Max Planck Encyclopedia of Public International Law*. Available online:

<https://opil.ouplaw.com/display/10.1093/law:epil/9780199231690/law-9780199231690-e2134#> (accessed: 19th January 2026).

—, 2019, ‘Robot, Inc.: Personhood for Autonomous Systems?’, 88 *Fordham L Rev* 591

Wall, S., 2013a, ‘Enforcing Morality’, *Criminal Law and Philosophy* 7: 455-471

——, 2013b, ‘Moral Environmentalism’ in C. Coons and M. Weber, *Paternalism: In Theory and Practice*, Cambridge: Cambridge University Press

——, 2021, ‘Perfectionism in Moral and Political Philosophy,’ *The Stanford Encyclopedia of Philosophy* (Fall 2021 Edition), E. N. Zalta (ed.) Available online: <https://plato.stanford.edu/archives/fall2021/entries/perfectionism-moral> (accessed: 19th January 2026).

Wallach, W. and Allen, C., 2009, *Moral Machines: Teaching Robots Right from Wrong*, Oxford University Press, Oxford

Walzer, M., 1977, *Just and unjust wars : a moral argument with historical illustrations*, 1st edition, New York: Basic Books.

Wang F., Kaushal R., Khullar D., 2019, ‘Should health care demand interpretable artificial intelligence or accept ‘black box’ medicine?’ *Ann Intern Med* 59–61. Available online: <https://doi.org/10.7326/M19-2548> (accessed: 19th January 2026).

Watson, G., 2001, ‘Reason and Responsibility’, *Ethics* 111: 374-394

Watson, L., 2007, ‘Pornography and Public Reason’, *Social Theory and Practice*, Vol. 33, No. 3.

Watson, D. and Floridi, L., 2020, ‘The Explanation Game: A Formal Framework for Interpretable Machine Learning’, *Synthese* 198: 9211–9242.

Watson, L., and C. Hartley, 2018, *Equal Citizenship and Public Reason: A Feminist Political Liberalism*, New York: Oxford University Press

Wellman, C., 1985, *A Theory of Rights: Persons under Laws, Institutions, and Morals*, Totowa, NJ: Rowman & Allanheld

———, 1995, *Real Rights*, New York: Oxford University Press

Wellman, C., 2014, 'Procedural Rights', *Legal Theory*, 20(4): 286–306

Whittaker, S., 2005, *Liability for Products: English Law, French Law, and European Harmonization* (Oxford, 2005; online edn, Oxford Academic, 22 Mar. 2012). Available online: <https://doi.org/10.1093/acprof:oso/9780198256137.001.0001> (accessed 20th January 2026).

Wieland, J.W., 2017, 'Introduction: The Epistemic Condition', in P. Robichaud, and J.W. Wieland (eds), *Responsibility: The Epistemic Condition* (Oxford, 2017; online edn, Oxford Academic, 20 July 2017). Available online: <https://doi-org.ezproxy-prd.bodleian.ox.ac.uk/10.1093/oso/9780198779667.003.0017>. (accessed: 7th January 2026)

Wenar, L., 2005, 'The Nature of Rights', *Philosophy & Public Affairs*, 33(3): 223–252

———, 2005[2021], 'Rights', *The Stanford Encyclopedia of Philosophy* (Spring 2021 Edition), E. N. Zalta (ed.)

West, C., 2022, 'Pornography and Censorship', *The Stanford Encyclopedia of Philosophy* (Winter 2022 Edition), E. N. Zalta and U. Nodelman (eds.)

Wilder, W., 2024, 'Virtual Vice: The Urgent Need to Reassess Ashcroft v. Free Speech Coalition in the Age of AI-Generated Child Pornography', *Law & Psychology Review* 48: 169-196. HeinOnline.

Williams, B., 1981, *Moral Luck: Philosophical Papers 1973–1980*. Cambridge: Cambridge University Press.

Williams, G., 1957, ‘Vicarious Liability and the Master’s Indemnity’, *The Modern Law Review*, 20(3):220–235. Available online: <http://www.jstor.org/stable/1090607> (accessed: 20th January 2026).

Wise, S. M., 1998, ‘Hardly a Revolution—The Eligibility of Nonhuman Animals for Dignity-Rights in a Liberal Democracy’, *Vermont Law Review* 22: 793

——, 2010, ‘Legal Personhood and the Nonhuman Rights Project’, *Animal Law Review* 17(1)

Wolf, S., 2001, ‘The Moral of Moral Luck,’ *Philosophic Exchange* 31: 4–19.

Wolff, J., 2024, ‘Introduction’ in Mantouvalou, V. and Wolff, J. (eds.), 2024, *Structural Injustice and the Law*, UCL Press.

Wong, M., 2026, ‘A Tipping Point in Online Child Abuse’, *The Atlantic*. Available online: <https://www.theatlantic.com/technology/2026/01/ais-child-porn-problem-getting-much-worse/685641/> (accessed: 24th January 2026).

Wood, T., Deep AI. Available online: <https://deepai.org/machine-learning-glossary-and-terms/unsupervised-learning> (accessed: 19th January 2026).

Wright, P. J., Tokunaga, R. S., Kraus, A., 2016, ‘A meta-analysis of pornography consumption and actual acts of sexual aggression in general population studies’, *J. Commun* 66(1): 183-205

Wu, S., 2013, ‘Risk Management in Commercializing Robotics’, 6-8 (Apr. 3, 2013).

——, 2016, ‘Product Liability Issues in the U.S. and Associated Risk Management’ in M. Maurer *et al.* (eds.) *Autonomous Driving. Technical Legal and Social Aspects*, SpringerOpen.

—, 2020, ‘Autonomous vehicles, trolley problems, and the law’, *Ethics and Information Technology* 22(1):1-13.

Wyre, R., 1992, ‘Pornography and Sexual Violence: Working with Sex Offenders’, in C. Itzin (ed.) *Pornography: Women, Violence and Civil Liberties*, Oxford: Oxford University Press

Young, F., 2025, ‘AI as a Product: The Next Frontier in Product Liability Law’, *University of Illinois Chicago* 13 October 2025. Available online: https://library.law.uic.edu/news-stories/ai-as-a-product-the-next-frontier-in-product-liability-law/#_ednref17 (accessed: 20th January 2026).

Young, I., 2004, ‘Responsibility and Global Labour Justice’, *The Journal of Political Philosophy* 12(4): 365-388

—, 2006, ‘Responsibility and Global Justice: A Social Connection Model’, *Social Philosophy and Policy* 23(1): 102-130

—, 2011, *Responsibility for Justice*, Oxford: Oxford University Press.

Yuracko, K., 2003, *Perfectionism and Contemporary Feminist Values*, Bloomington: Indiana University Press

Zelizer, V., 1996, *The Social Meaning of Money*, Princeton: Princeton University Press

Zerilli, J., Knott, A., Maclaurin, J., Gavaghan, C., 2019, ‘Transparency in Algorithmic and Human Decision-Making: Is There a Double Standard?’ in *Philosophy & Technology* 32: 661-683

Zhang, E., 2021, ‘Why Sex? Sex-Bots from a Daoist Perspective’, in R. Fan and M. Cherry (eds), *Sex Robots: Social Impact and the Future of Human Relations*, Springer

Zheng, R., 2019, 'What kind of responsibility do we have for fighting injustice? A moral-theoretic perspective on the social connections model', *Critical Horizons* 20(2): 109-126

Zhou, Y. and Paul, B., 2016, 'Lotus Blossom or Dragon Lady: A Content Analysis of 'Asian Women' Online Pornography', *Sexuality and Culture*, 20: 1083–100

Zimmerman, M. J., 1985, 'Intervening Agents and Moral Responsibility', *The Philosophical Quarterly* 35(141):347.

—, 1986, 'Negligence and Moral Responsibility', *Noûs*, 20(2): 199–218.

Zwolinski, M., Ferguson, B. and Wertheimer, A., 2022, 'Exploitation', *The Stanford Encyclopedia of Philosophy* (Winter 2022 Edition), E. N. Zalta & U. Nodelman (eds.). Available online: <https://plato.stanford.edu/archives/win2022/entries/exploitation/> (accessed: 20th January 2026).

Case law

Anderson v. General Motors [1985] Civ. 21876 Court of Appeal, Third District, California

Armes v Nottinghamshire County Council [2017] UKSC 60

Ashcroft v. Free Speech Coalition 535 U.S. 234, 241 (2002).

Barclays Bank v Various Claimants [2020] UKSC 13

Bazley v Curry [1999] 2 SCR 534

Bellman v Northampton Recruitment Ltd [2018] EWCA Civ 2214, [2019] ICR 459

Bemba Gombo No. I [2009], Decision Pursuant to Art. 61(7)(a) and (b) of the Rome Statute on the Charges of the Prosecutor Against Jean-Pierre Bemba Gombo, Bemba Gombo (ICC-01/05-01/08), Pre-Trial Chamber II, 15 June 2009

Bemba Gombo No.2 [2016] Judgment pursuant to Art. 74 of the Statute, Bemba Gombo (ICC-01/05-01/08), Trial, Chamber III, 21 March 2016

Bolger v. Amazon LLC 53 Cal.App.5th 431 (2020).

Burwell v. Hobby Lobby Stores, Inc. [2014] 573 U.S. 682

Caparo Industries Plc v Dickman [1990] 2 AC 605

Carlill v. Carbolic Smoke Ball Company [1892] EWCA Civ 1.

Catholic Child Welfare Society (CCWS) [2012] UKSC 56, [2013] 2 AC 1

Citizens United v. Federal Election Commission [2010] 558 U.S. 310

Cox v Ministry of Justice [2016] UKSC 10

Dartmouth College v. Woodward [1819] 17 US 518

Dean & Chapter Of Rochester Cathedral v Leonard Debell [2016] EWCA Civ 1094 CA (Civ Div)

Donoghue v Stevenson [1932] AC 562

DPP v Majewski [1976] UKHL 2

Dubai Aluminium Co Ltd v Salaam [2002] UKHL 48

Dudgeon v UK [1981] ECHR 22 Oct 1981

EB v Order of the Oblates of Mary Immaculate in the Province of British Columbia [2005] 3 SCR 45

Escola v. Coca Cola [1944] 150 P.2d 436

Garcia v Character.ai [2025] Case No. 6:24-CV-01903-ACC-UAM

Greenman v. Yuba Power Products, Inc. [1963] 59 Cal.2d 57

Griswold v. Connecticut [1965] 381 U.S. 479

Hadley v Baxendale [1854] EWHC J70

Heaven v Pender (1883) 11 QBD 503

Henningsen v. Bloomfield Motors [1960] 32 N.J. 358.

Hollis v Vabu (2001) 207 CLR 21

Honeywill and Stein Ltd v Larkin Brothers Ltd [1934] 1 KB 191 (CA) 196

Jacobi v Griffiths [1999] 2 SCR 570

JGE v English Province of Our Lady of Charity [2011] EWHC 2871 (QB), [2012] 2 WLR 709

JGE v English Province of Our Lady of Charity [2012] EWCA Civ 938

John Doe v Bennett [2004] 1 SCR 436

Jolley v Sutton London Borough Council [2000] 1 WLR 1082

Lawrence v. Texas [2003] 539 U.S. 558

Levitt v Euro Building and Maintenance Contractors Ltd [2019] EWHC 2926 (QB).

Lister v Hesley Hall Ltd [2001] UKHL 22, [2002] 1 AC 215

Loomis v. Amazon.com LLC 63 Cal.App.5th 466 (2021)

Majrowski v Guy's & St Thomas's NHS Trust [2005] IRLR 340

Masterpiece Cakeshop, Ltd. v. Colorado Civil Rights Commission [2018], 584 US (2018)

Miller v R [2010] EWCA Crim 1578

Modinos v Cyprus [1993] 16 EHRR 485

Mohamud v Morrison Supermarkets plc [2016] UKSC 11

Ng Huat Seng v Munib Mohammad Madni [2017] SGCA 58, [2017] 2 SLR 1074

Nonhuman Rights Project v. Lavery [2014], *People ex rel. Nonhuman Rights Project, Inc. v Lavery* [2014] 124 A.D.3d 148, 998 N.Y.S.2d 248

Norris v Ireland [1988] 6/1987/129/180

Prosecutor v. Brima [2006] (scsl- 04-16-t) scsl 35 (31 March 2006)

Prosecutor v. Delalić [1998] Case No. IT-96-21- T, Trial Chamber Judgment, ¶¶ 437, 439 (Int'l Crim. Trib. for the Former Yugoslavia Nov. 16, 1998). Available online: http://www.icty.org/x/cases/mucic/tjug/en/981116_judg_en.pdf (accessed: 11th January 2026).

Prosecutor v. Limaj et al. [2005] Case No. it- 03- 66- t, Trial Chamber Judgment (International Criminal Tribunal for the Former Yugoslavia 30 November 2005).

Prosecutor v. Nahimana [2007] Case No. ictr 99-52-a, Appeals Chamber, Judgment (International Criminal Tribunal for Rwanda 28 November 2007).

Prosecutor v. Stanislav Galić [2003] Case No.it-98-29-t, Trial Chamber Judgment (International Criminal Tribunal for the Former Yugoslavia 5 December 2003).

Prosecutor v. Tadić [1999] Case No.it-94-1-a, Appeals Chamber Judgment, (International Criminal Tribunal for the Former Yugoslavia 15 July 1999).

R v Brown [1993] UKHL 19, 1 AC 212

R v Heard [2007] EWCA Crim 125

R v Lipman 1 QB 152

Reliable Cons. v. Earle [2008] 517 F.3d 738 (5th Cir. 2008)

Reeves v Commissioner of Police of the Metropolis [1999] 3 WLR 363

Robinson v Chief Constable of West Yorkshire Police [2018] AC 736

Roe v. Wade [1972] 410 US 113

Roman Catholic Episcopal Corporation of St George's v John Doe (a pseudonym) and John Doe (a pseudonym) [2004] 1 SCR 436

Rushbond v JS Design [2021] EWCA Civ 1889

Santa Clara Co v. Southern Pacific Railroad [1886] 118 US 394

Scott v Davis (2000) 204 CLR 333

Sindell v Abbott Laboratories 26 Cal. 3d 588

Skandinaviska Enskilda Banken AB v Asia Pacific Breweries (Singapore) Pte Ltd [2011] SGCA 22, [2011] 3

Stansbie v Troman [1948] 2 KB 48

State v. Horton [1905] 51 S.E. 945

The Wagon Mound (No.1) [1961] AC 388

Toonen v Australia [1994] Communication No.488/1992, 4/4/1994: Australia

Trotman v North Yorkshire County Council [1999] LGR 584

Trustees of Dartmouth College v Woodward [1819] 17 U.S. 518

Twyman v. Twyman [1993] 855 S.W.2d 619

Victoria Laundry (Windsor) Ltd v Newman Industries Ltd [1949] 2 KB 528

Wm Morrison Supermarkets plc v Various Claimants [2020] UKSC 12, [2020] AC 989