

Justifying Constraint in the Legal Regulation of Reproduction

A thesis submitted to the University of Manchester for the degree of Doctorate in
Bioethics and Medical Jurisprudence in the Faculty of Humanities.

2016

Samuel Edward Patrick Walker

School of Law/Centre for Social Ethics and Policy

Word count: 82492

Thesis Contents

<u>Abstract</u>	<u>3</u>
<u>Declaration</u>	<u>4</u>
<u>Copyright Statement</u>	<u>4</u>
<u>Dedication</u>	<u>5</u>
<u>The Author</u>	<u>6</u>
<u>Introduction</u>	<u>7</u>
<u>Structure of the Thesis</u>	<u>48</u>
<u>Part 1</u>	
<u>Chapter 1: Can Morality Justify Population-wide Constraint?</u>	<u>50</u>
<u>Chapter 2: Can Rational Constraint Secure Co-operation?</u>	<u>86</u>
<u>Chapter 3: Is a Waiver of Human Rights a more Rational and Beneficial Theory of Rights?</u>	<u>122</u>
<u>Part 2</u>	
<u>Chapter 4: Potential Persons and the Welfare of the (Potential) Child Test</u>	<u>154</u>
<u>Chapter 4: Subsection: Additional Comments on Impersonalism</u>	<u>173</u>
<u>Chapter 5: Applying the Actual/Potential Person Distinction to Reproductive Torts</u>	<u>190</u>
<u>Part 3</u>	
<u>Open questions for a theory of rational constraint</u>	<u>215</u>
<u>Thesis Conclusion</u>	<u>220</u>
<u>Thesis Bibliography</u>	<u>226</u>
<u>Appendix</u>	<u>234</u>

Abstract

The University of Manchester, Samuel Edward Patrick Walker, PhD in Bioethics and
Medical Jurisprudence

Thesis Title: Justifying Constraint in the Legal Regulation of Reproduction

2015

This thesis seeks provide an original contribution by extending rational choice theory into a general theory of law which has not been done before. I term this theoretical framework *rational constraint* to distinguish it from other rational choice and contractarian theories. This is predicated on the increasing heterogeneity of contemporary populations and the claim that moral claims cannot resolve conflicts because they are not truth-apt (as I argue in chapter 1). I seek to extend rational choice theory by further developing the tradition of social contract theory as it applies to law in the contractarian tradition of Thomas Hobbes and David Gauthier. This can roughly be termed the contractarian version of social contract theory (in the introduction I distinguish this tradition from the other social contract tradition of contractualism). This tradition takes rationality to be a practical method for determining action based upon self-interest – this assumes that agents are not concerned about the wellbeing of others. Even so a broad range of restrictions are possible – this thesis seeks to take Gauthier’s theory (as the most contemporary and developed contractarian theory) further by providing a system that takes account of higher-order constraints as well. This approach is not concerned with the application of different competing sets of moral claims – rather the application of self-interested rationality to law is the focus and the original contribution of this thesis. Ultimately, I seek to provide a method for designing legal rules that can minimise conflict and cost in a heterogeneous population. The subject that I apply this framework to is reproduction which is non-economic in nature thus extending rational choice beyond its normal economic haunts. Moreover it is an area of law that concerns a part of life subject to a great deal of moral controversy thus demonstrating the superiority of the extended rational choice framework over moral systems in designing laws.

Declaration

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright Statement

- i. The author of this thesis (including any appendices and/or schedules to this thesis) owns any copyright in it (the “Copyright”) and s/he has given The University of Manchester the right to use such Copyright for any administrative, promotional, educational and/or teaching purposes.

- ii. Copies of this thesis, either in full or in extracts, may be made only in accordance with the regulations of the John Rylands University Library of Manchester. Details of these regulations may be obtained from the Librarian. This page must form part of any such copies made.

- iii. The ownership of any patents, designs, trade marks and any and all other intellectual property rights except for the Copyright (the “Intellectual Property Rights”) and any reproductions of copyright works, for example graphs and tables (“Reproductions”), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property Rights and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property Rights and/or Reproductions.

- iv. Further information on the conditions under which disclosure, publication and exploitation of this thesis, the Copyright and any Intellectual Property Rights and/or Reproductions described in it may take place is available from the Head of School of Law (or the Vice-President) and the Dean of the Faculty of Life Sciences, for Faculty of Life Sciences’ candidates.

Dedication

I dedicate this thesis to my family and friends. Special thanks go to my parents, Phil and Elayne Walker, for all the support and encouragement throughout the PhD process, and to my brother, Luke Walker, for always keeping my feet on the ground. Thanks also go to my grandparents for their support of my PhD studies. I must also thank my friends Sophie Bartsch and John Coggon who have put up with my complaints over the past few years and for their counsel to just get on with it. Finally, I would like to thank my supervisor Professor Søren Holm for his comments, feedback and advice during the course of my PhD. I would also like to thank Dr Alex Mullock for her help in preparing this thesis.

The Author

I undertook my first degree at the University of Aberystwyth in Law from which I graduated in 2008. During the course of my time at Aberystwyth undertook a medical ethics and law course which introduced to me the exciting area of law, society, ethics and medicine. This initial introduction has become a passion for me and prompted me to continue my studies in this highly relevant and cutting edge area of law. Consequently, I studied a Masters in Public Law at the University Of Auckland in New Zealand. The major part of my studies there was engaged in a thesis on euthanasia in cases where the patient cannot communicate with anyone, in the main due to severe brain trauma. This thesis considered non-voluntary euthanasia and how to clarify the law in this area. During my Masters I decided to embark upon a PhD and thus applied for the Bioethics and Medical Jurisprudence PhD programme at the University of Manchester. This PhD programme interested me both for its focus on bioethics and law, and because of the great community of researchers and academics there. After being encouraged to undertake a new area of research I decided to pursue research in the area of reproduction and biotechnologies. My time in Manchester has been challenging and enjoyable in equal measure.

Introduction

The Problem

The dominance of moral systems in particular geographic regions is declining in the developed world (the decline in the dominance of Christianity in the UK is striking example of this as well as the increasing ethnic diversity of contemporary nations). This coincides with hugely increased material wealth and the rapid expansion of the population leading to a population of agents that is heterogeneous, materially wealthy and, while interdependent, largely anonymous to each other. These two predicates (the decline of moral dominance and the heterogeneity of populations) lead me to seek an alternative means of designing legal systems that can take account of and accommodate the increasing differences between individuals. By accommodate I mean a system of law that imposes the minimal costs needed to maintain social cohesion (that is co-operation) while leaving agents to act as freely as possible. Such a system is desirable because it requires the least amount of social resources to enforce as each individual has an incentive to comply with the system whilst simultaneously preventing conflicts based upon moral claims.

Consequently, this thesis seeks to perform two tasks. First, to demonstrate that morality cannot justify constraint for a population of heterogeneous agents because such justification requires metaphysically queer facts. This leaves rational constraint as the only justification for population wide constraint. Rational constraint is constraint on an agent's conduct that confers benefits to those who behave according to constraint making possible co-operation thus agents can receive greater benefits than would otherwise be the case. Rational constraint is based upon the methodology of rational choice within a strategic context – thus constraint is rationally justified when this leads to higher payoffs for participants than alternative actions they can take given the possible actions of others. Second, to use the theory of rational constraint and the rational choice methodology to develop a new theory of law that can critique current legal rules and outline alternatives aimed at increasing the scope of individual preference-maximisation while limiting the role of the law to interpersonal conduct. This means a rejection of moral claims to justify colonising the preferences held by agents and enforcing a particular moral system. Instead, the following form of legal analysis uses practical rationality to determine the validity of legal regulation and identify unnecessary constraints applied to individuals whilst

justifying and enforcing those constraints needed for maintaining the overall co-operative system. This thesis attempts to develop a general theory of law based upon the contractarian tradition of social contract theory and following in the footsteps of Thomas Hobbes and David Gauthier. I aim to produce a theoretical framework containing a method for designing legal rules without any moral underpinnings and which can regulate the conduct of agents in a heterogeneous population.

These two tasks are important because of the increasing heterogeneity of contemporary populations that possess divergent moral attitudes which cannot be resolved through moral discourse because moral claims are not truth-apt. One need only observe the change in attitudes in contemporary ‘Western’¹ society to see the increasing divergence within populations. Competing moral systems now exist alongside each other and ‘Western’ morality is no longer the monoculture it once was (or seemed to be). The 20th and 21st centuries have seen a sharp decline in the power of the dominant moral narrative to induce behavioural conformity.

In the United Kingdom, for example, this has primarily revolved around the declining power of Christianity to enforce behavioural norms by requiring that all other constraints comply with its ordinances. One study shows that church members constitute only 10% of the UK population² and the 2011 UK Census showed that 26.13% of the population classify themselves as having no religion³ (a considerable increase from the 2001 UK Census when 15.5% identified themselves as having no religion⁴). This suggests that the power of Christianity and religion (even if all religious people belonged to the same religious sect) would not provide over a quarter of the population with a reason to comply with its ordinances.

Religious belief is just one example of a set of prescriptions based on mind-independent moral properties, other non-religious sets of moral systems (such as Utilitarianism, deontology, Kantian ethics and so on) also require, permit or forbid different behaviours. These theories can all be rejected (as we shall see) on the basis

¹ The concept of western is problematic because it is a nebulous term. By western I mean the liberal capitalist model of democracy and pluralism found in Europe and North America. This is in contrast to more moral-authoritarian parts of the world such as Saudi Arabia with its strict religious law or China with its autocratic communist dogma.

² Ruth Gledhill, ‘Church attendance has been propped up by immigrants, says study’ *The Guardian* (London, 3 June 2014) <http://www.theguardian.com/world/2014/jun/03/church-attendance-propped-immigrants-study> accessed 17/11/2015.

³ Office of National Statistics ‘2011 Census’ at <http://www.ons.gov.uk/ons/guide-method/census/2011/index.html> accessed 17/11/2015.

⁴ Office of National Statistics ‘2001 Census’ at www.ons.gov.uk/ons/guide-method/census/census-2001/index.html accessed 17/11/2015.

that they require belief in mind-independent moral claims which comes to down to the problem of moral foundationalism and the problem of explaining how a *moral* claim can be anything other than a mind-dependent creation of an agent. In turn, the more closely constraints rely on or contain moral artefacts the less justified those rules are likely to be because they will only represent the beliefs of one particular moral system. Investigating the basis for moral foundationalism – even with the intention of strengthening the role of morality – it is not long before the nature of moral claims comes into question.

Consequently, in a population of heterogeneous agents, constraint must be in their self-interest for no sub-group of that population can justifiably impose its moral system on others. Moreover, the cost of maintaining such tyrannical constraint is considerable because agents would expend resources to evade the system – think of the speakeasies of the Prohibition-era United States of America – such resources could be better spent on the projects of agents rather than conflict between them. The purpose of this thesis is to present an alternative method for analysing, critiquing and reforming legal rules which would secure co-operation whilst permitting agents as much freedom as possible.

This introduction begins by explaining the need for such constraint and what can justify it. It will also state why it is important to address the legal regulation of reproduction and why it has been chosen as the subject of chapters three and four. The research questions that form the basis of each chapter will be outlined to contextualise each step that the chapters make. Once the key areas are identified more detail will be provided on the philosophical and legal theory that surrounds this project – this will situate the work carried out here within the wider academic field. Of particular note is the distinction between the theoretical approach taken here (rational constraint theory) with other moral and ethical theories and the links the resulting legal theory has with other jurisprudential theories. This will lead to some repetition but it should also become successively more detailed and clearer as we progress. Additionally, this introductory outline is recommended by the Bioethics and Medical Jurisprudence PhD program itself.

Philosophical and legal background

The purpose of this thesis is to explain why moral claims cannot resolve inter-agent conflict and explain the necessity for an alternative, to set out this alternative and explain how it would work, and to apply this approach to the law and explore the options that it presents to us. Here I present an outline of the core points of rational constraint – that is a non-moral rational choice general theory of law containing a method of practical rationality that can justify legal constraint and design specific legal rules.

An outline of rational constraint theory and this thesis

Morality fails because it must efface the differences between individual agents reducing them all too homogeneous moral subjects who respond to and behave according to a set of correct moral facts. But moral constraints, which are limitations on an agent's conduct justified by moral claims, cannot be justified because the type of moral claims needed to generate constraint must 'entail irreducibly normative reasons' in order to be binding upon *all* agents.⁵ John Mackie first articulated the idea of queerness in moral properties, stating that 'if there were objective values [real moral properties], then they would be entities or qualities or relations of a very strange sort utterly different from anything else in the universe'.⁶ This argument arises from the metaphysical queerness of moral properties because 'our concept of a moral fact is a concept of an *objectively prescriptive* fact' but 'there *are* no objectively prescriptive facts or properties'.⁷ Yet the 'main tradition of European moral philosophy includes the contrary claim'.⁸

Nevertheless, it seems that opponents of moral error theory can 'maintain that it is a fundamental fact about reality that there are irreducibly normative reason relations' by simply claiming that irreducible normativity is *not queer*.⁹ If the objective prescriptivity of moral claims relies on the notion of irreducible normativity and if this can only be asserted, irreducible normativity cannot be the basis of constraint for all agents *because some agents will reject it*. In contrast, non-cognitivist theories hold that moral claims are an expression or projection of our sentiments,

⁵ Jonas Olson, *Moral Error Theory* (Oxford University Press 2014) 117.

⁶ John Mackie, *Ethics: Inventing Right and Wrong* (Penguin Books 1990) 38.

⁷ Alexander Miller, *Contemporary Metaethics* (Polity Press 2013) 105.

⁸ Mackie 30.

⁹ Olson 136.

emotions or attitudes onto the world. As Simon Blackburn eloquently puts it, ‘[p]rojecting is what Hume referred to when he talks of “gilding and staining all natural objects with the colours borrowed from internal sentiment”, or of the mind “spreading itself on the world”’.¹⁰ Here, moral claims remain but they are relativistic to the individual expressing that claim, thus they cannot provide a justification for constraint amongst all agents because they are idiosyncratic to the agent expressing the moral claim rather than a universal truth. Of course it may be that there are objective mind-independent moral properties but the arguments within cognitivist theories and the arguments advanced by moral error theorists suggest that there are significant reasons to doubt the existence of mind-independent moral claims – even without introducing the problem of how we can epistemologically identify such moral claims.

Consequently, moral claims cannot justify constraint to those who do not already accept that particular moral system. Hence, in a *homogeneous* population law and morality can reflect each other to a high degree – to the point that they could be one and the same – but once differing moral claims come into play this close connection between law and morality cannot be maintained. Instead, constraint for the population must be justified by another framework and the framework proposed here is a (contractarian-derived) theory of rational constraint based upon the rational choice methodology. Rational constraint is based upon the following principle: agents benefit from the constraint and co-operation of others but they will only co-operate if all agents are constrained thus a system of constraint applied to all, benefits all and secures co-operation. Benefits mean a number of things – not having to invest resources to protect oneself from predation, a high degree of certainty in interaction with others reducing information costs, the creation of resources and engaging in activities only available from group co-ordination, a mechanism for redressing violation of constraint and the freedom from interference within the system of constraint to live the life agents desire. The role of benefit is crucial in ensuring voluntary compliance by agents which is the least costly method for maintaining co-operation.

¹⁰ Simon Blackburn, *Spreading the Word* (Oxford University Press 1984) 171.

David Gauthier's contractarian theory expounded in his book *Morals by Agreement*¹¹ is the start point of developing this theoretical framework. The reason for selecting Gauthier's theory is that it does not make any *moral* assumptions – the only assumptions Gauthier does make are that agents are practically rational, they form a set of considered formally rational preferences and that they do not have any concern for others. Human agents must be capable of means-end rationality in order to determine the most beneficial course of action. While people do not always act rationally they are, in principle, capable of it – if no one could act (even approximately) rationally then we would be precluded from using reasons as an argument for carrying out or prohibiting certain actions. This seems to be an unavoidable assumption about human agents but it is a reasonable one given that people do seem to act rationally – a trivial example would be someone deciding to cook lasagne on a Saturday and buying the ingredients for it beforehand.

Thus, the system of constraint 'is a system of a particular sort of constraints on conduct – one whose central task is to protect the interests of persons'.¹² There are two points relating to this theory which must be noted: first, it may not be the case that all agents form a single group for the purposes of constraint and, second, it *will* be rational for a few agents to violate the system of constraint which can be tolerated as long as a critical number of agents abide by the system. The first of these issues raises questions regarding inclusion while the second raises the problems of compliance – the fact that rationality may require defection from compliance and the problem of specifying the relevant population for the purposes of interaction, constraint and co-operation. A skete monk certainly accepts and complies with some constraints on his behaviour when interacting with his brother monks but we would be hard pressed to say they are part of the constraints of people on the other side of the world given the self-sufficiency of their population. Population, in the sense of groups of interacting agents, has a significant impact on who interacts with whom. This shows that we cannot assume constraint is universal because agents can isolate themselves from interaction – in the skete monastic order monks retain some community but a hermit maintains no community at all. Monastic communities are unusual as membership is

¹¹ David Gauthier, *Morals by Agreement* (Clarendon Press 2001).

¹² Mackie (n6) 106: He calls this 'morality in the narrow sense', though it is misleading to continue to use *moral* in a post-moral error theory discourse.

voluntary, while in most populations of agents membership is by birth (although some become members through naturalisation of nationality).

Yet, in the contemporary globally integrated world there are very few instances of isolated interaction – thus, boycotting a particular company or product in one country *will* affect the lives of agents in other countries. The example of a skete monk was chosen because it illustrates the claim that he is only constrained in relation to his brother monks *because those are the people with whom he interacts*. This is not a perfect example as even a ‘hermitic life in a largely empty country such as Mongolia’ is still lived within ‘Mongolia, a recognised and discrete political entity’.¹³ One result of a theory of rational constraint based upon interaction is that if an agent can avoid interaction then he also avoids constraint. When discussing issues we must be careful to identify the population of which we are speaking – in the case of legal issues this is simpler because the concept of a particular jurisdiction is (normally) quite concretely defined. We shall see, however, in chapter four that some argue for the inclusion of potential future agents but there is serious doubt about whether we can interact with (and thus benefit) future people.

The second issue relates to the free-rider problem in that a sufficiently large population can accommodate some agents defecting from the constraint system – indeed in some cases the actions of a few predator-strategies are rational for those few agents. The rationality of these strategies can be decreased through legal sanction – indeed, according to the theory set out here this is the primary justification of judicial systems – but some defection can be tolerated as long as defections remain low enough that agents *do not expect* defection by others more than they expect co-operation. Once the point is reached that predation/exploitation is expected with a greater frequency than co-operation, it will become rational for all agents to defect defeating the co-operative system. Thus, within a given population, some agents will be able to benefit from non-compliance and under a theory of rational constraint they *should* violate co-operation – this is why judicial systems are so important because they can modify the context in which interaction takes place by increasing the costs and reducing the benefits of predator-strategies. This is the key point for justifying both legal rules themselves and the critique of judicial systems.

¹³ Iain Brassington, *Truth and Normativity* (Ashgate Publishing Ltd 2007) 18: Brassington’s discussion takes place in the context of the notion of *consent* to being governed by the state in which one finds oneself.

As already intimated, one of the most powerful tools for inducing conformity is a legal system enforced by investigatory agencies and the application of sanctions to non-conforming individuals. It is important to remember that legal systems are one form of behavioural constraint situated within a broader socio-political arena containing other systems of constraint – the reciprocal influence of each system on the others deserves analysis in itself but that is, unfortunately, beyond the scope of this thesis. Yet, laws can be onerous and unjustified – for example, the welfare of the child test (found in the Human Fertilisation and Embryology Act 1990 as amended by the 2008 Human Fertilisation and Embryology Act) imposed on infertile couples is not necessary for ensuring co-operation yet they allow others to inveigle themselves into the decision making process. Additionally, the rules themselves can be used as means of imposing particular normative moral judgements upon agents. In the case of reproduction – up until the 2008 amendment – the Human Fertilisation and Embryology Act required that ‘the need of that child for a father’ be satisfied before assistance would be given thus (re)enforcing the hetero-normative model of the family.¹⁴ The revised version requires ‘supportive parenting’ which leaves open the possibility of value judgements about what sort of families there should be or who would make a good parent whilst simultaneously expanding parenting beyond the hetero-normative male-female dyad.¹⁵

The theory of rational constraint determines which laws are unjustified (because they are *irrational* or because they are not the *most rational*) and, it might be argued, this would justify prospective parents failing to comply with these constraints (i.e. by acting deceptively or by using services acquired abroad). However, the issue of non-compliance is wide enough to encompass anyone who does not comply – including murderers, rapists and assaulters. We must, therefore, be able distinguish between those individuals and the infertile couple beyond the mere fact of non-compliance. This calls for a justification of constraint that can explain why some behavioural limitations are desirable, why some are best achieved using legal rules and why constraints should be ignored or obeyed which, taken together, will justify the use of coercive measures to prevent co-operation damaging non-compliance. This calls for a complete political and legal theory of rational constraint which would be able to justify constraint between heterogeneous agents.

¹⁴ Human Fertilisation and Embryology Act 1990 s.13(5) [pre-amendment].

¹⁵ Human Fertilisation and Embryology Act 2008 s.14(2).

By developing a non-moral justification for legal rules I hope to identify alternative legal rules that can mediate between the different moral claims that agents make. To be clear, whilst the law cannot be based upon moral claims it can and must take account of the moral values of agents when determining appropriate legal rules. Moral claims cannot, however, be more than the attitudes and sentiments of an agent – consequently legal systems should treat moral values as part of an agent’s preference set rather than as an objective fact the world. Agents can make moral claims about how *they want to behave* but they cannot justify imposing those behavioural constraints upon others *on moral grounds*. Moral error theory shows that moral systems cannot fulfil their justificatory obligations for constraint. Agents *can*, however, make claims about how others should behave on the basis of rational constraint.

Thus a rationalist method for critiquing legal rules and systems would fulfil two functions. First, it would allow us to consider moral values without using them as a general foundation for constraint and, specifically, for the law. Second, alternative legal structures can be identified and compared to the existing system through the effect on agents. On this theory, a system of constraints is not in place to promote *right* and prohibit *wrong* conduct. Systems of constraint are there to mediate between conflicting agents and ensure compliance with a set of behavioural limitations that are *necessary* for agents to co-operate. As moral systems cease to dominate and heterogeneity increases the law must be purged of moral relics and we must prevent moral claims (whether new or old) becoming the basis for new laws. Interaction problems arise because of the heterogeneity of agents and their different conflicting goals. Consider Richard Braithwaite’s illuminating exemplar of interaction problems. Luke and Matthew live in two flats of a converted house:

Luke can hear everything louder than a conversation that takes place in Matthew’s flat, and vice versa ... [imagine] each of them has only the hour from 9 to 10 in the evening for recreation, and that it is impossible for either to change to another time. Luke’s form of recreation is to play classical music on the piano for an hour at a time, and that Matthew’s amusement is to improvise jazz on the trumpet for an hour at once. ... Suppose that the satisfaction each derives from playing his instrument for

the hour is affected, one way or the other, by whether or not the other is also playing.¹⁶

Here Braithwaite posits that ‘Luke’s first preference is for him alone to play’ and – while Matthew is indifferent between both playing and himself playing alone – he dislikes Luke playing alone.¹⁷ Consequently they have conflicting preferences for the use of the one hour in the evening and need to adopt behavioural constraints that minimises the conflict between them. This role is played by legal regulation and, although not every constraint should be a *legal constraint*, all legal constraints should be judged by the metric of rational constraint.

What are ‘agents’?

The conception of an agent is crucial to the scope of this theory – I deal with this before describing the distinction between the rational and the moral. A crucial feature of this approach is that only agents are relevant to the system of constraint because only these kinds of entities can constrain their conduct. Only agents can engage in strategic interaction where they can choose certain actions aware that these choices will affect how others behave. Implicitly, an agent must be a self-conscious entity able to choose between the different options available to them. Agents play the central role in determining acceptable constraints because they will only engage in strategic interaction if it is beneficial for them. Thus, unlike other theories of law or moral, agents are not passive recipients of constraint rather they constitute those constraints through their participation in them. Only *persons* can participate in a system of constraints while non-persons can only be passive beneficiaries of that system.

Agents must be aware of themselves as distinct diachronic entities (that is entities which exist at two different points in. and thus over, time) who can modify their behaviour during repeated interaction with other similar entities. Agents must also be capable of engaging in practical rationality when reasoning through the different options available to them. This does not require that agents are rational in the substantive sense of rational objectives (if such rationality is possible), for human agents are not, but it does require that agents can think rationally in the sense of

¹⁶ Richard Braithwaite, *Theory of Games as a tool for the Moral Philosopher* (Cambridge University Press 1955) 8.

¹⁷ *Ibid* 9.

means-end reasoning. An agent must be able to work out the practical steps they need to take in order to achieve their goals. Agency is important for any theory of human behaviour but it is at the heart of constraint theory because only agents are capable of voluntarily constraining their actions. (Voluntary constraint is preferable to externally imposed constraint because it costs, in the sense of resources, less to maintain.)

The ‘importance of ... having a theory of co-personality [i.e. diachronic persons] stems from the four features of human existence’.¹⁸ These four features are as follows: first, that ‘our concern for our own future states is different ... from our concern for the future states of others’; second, ‘the phenomenon of survival, our continuing to be the persons we are now’; and (third and fourth) the ‘social phenomena of compensation and responsibility’.¹⁹ These features require diachronic agency because if ‘a person does an action, it is the same person who can later be held responsible for the action, and whom it is appropriate to punish or reward for doing’.²⁰ ‘In this personal identity is founded all right justice of reward and punishment’.²¹

A confusing profusion of terms has now arisen – person, co-personality and personal identity²². To an extent these terms are interchangeable – we might add ‘diachronic agent’ to this list – but they all refer to the same core concept: that agents are self-aware diachronic entities to whom actions can be assigned and who can voluntarily constrain their conduct by complying with a system of constraint. For the purposes of this thesis I will use the term ‘agent’ to denote diachronic persons because this term foregrounds the role of choice in the functioning of constraint systems.

An agent is, therefore, a psychological entity – albeit one that must be embodied. Personhood theory attempts to explain the concept *person* but for this thesis it is the psychological features of agents which are of interest. The crucial requirements are the capacities of self-consciousness and self-reflection through which an agent can analyse their behaviour and the behaviour of others altering their behaviour according to the circumstances. It also requires that agents can imagine alternative courses of actions for themselves and for those with whom they interact. An agent must possess

¹⁸ Marc Slors, *The Diachronic Mind* (Kluwer Academic Publishers 2001) 9.

¹⁹ *Ibid* 9.

²⁰ Sydney Shoemaker and Richard Swinburne, *Personal Identity* (Blackwell 1985) 71.

²¹ John Locke, *An Essay Concerning Human Understanding* (first published 1689, Everyman 1996) 186.

²² Personal identity is distinct from the other terms because it refers to a particular individual whereas the other terms refer to a class of entities – although personal identity necessarily implies the presence of personhood or co-personality.

the ‘concept of a self as a continuing subject of experiences and other mental states, and believes that it is itself such a continuing entity’.²³ We must have a self-conscious²⁴ awareness of ourselves and a sense of time²⁵ (of past, present and future) if we to be able to assess our conduct with each other. If year-ago-Sam were not present-day-Sam then we would not be able to speak of Sam as a unified entity over time that is responsible for past action. Diachronic unity is an essential component for constraint to be an intelligible concept.

Here we risk entering the dangerous territory of the constituting function of language. Briefly, the fact that we speak as though we are diachronic agents may constitute us *as* diachronic agents. Although we all recognise that our self-awareness is fragmented – we remember some events but not others, we behave differently in different contexts and play different social roles – our language and concepts are built on the concept of ourselves as persisting over time. We are trapped within a referential and linguistic framework that characterises us as the same entity throughout time. Nevertheless, the importance of agency in constraint theory does not stem from the linguistic devices that we use to speak of ourselves. Instead the importance of agency is based upon the capacity for self-modifying behaviour and the mental capacities needed to practically reason through alternative courses of action and their consequences. (It is possible that we are not diachronic unified entities but if this is the case then it would require the development of new concepts and language to cope with the absence of a subject persisting through time.)

These requirements of agency place emphasis on the psychological rather than the biological features of humans and the focus on mutually beneficial interaction combine to privilege the role of existing persons in constraint theory. We can only interact in a mutually beneficial manner with other agents who exist throughout the period of interaction. This also suggests that geography plays a role in determining with whom we interact – prior to the (re)discovery of the New World interaction between Americans and Europeans was impossible (although this lack of interaction was beneficial for the indigenous American population) – but globalisation has almost entirely ended the possibility of isolated communities that do not affect each other. Even so, the later discussions in this thesis will focus on England and Wales as a

²³ Michael Tooley, ‘Abortion and Infanticide’ (1972) 2 *Philosophy and Public Affairs* 37, 44.

²⁴ Joseph Fletcher, ‘Indicator of Humanhood: a tentative profile of man’ (1972) 2 *The Hastings Center Report* 1, 1.

²⁵ *Ibid* 2.

single socio-political jurisdiction – although comparisons will be made with other jurisdictions.

A sharp distinction between rationality and morality

This approach makes a sharp distinction between the moral and the rational. Morality makes metaphysical claims about the way the world is that agents should comply with. Rationality, in contrast, concerns reason for actions and the specific branch of rationality that is used in this theoretical framework is practical or instrumental rationality. Substantive rationality is an alternative account to practical rationality but substantive rationality should be rejected because it suggests that reason relates to external things lying in the world to which we can respond correctly or incorrectly. Substantive rationality is concerned with reason in the ‘standard normative sense’ in which we consider whether something ‘is a *good* reason – a consideration that really counts in favour of the thing in question’.²⁶ This is distinguishable from instrumental reason – when ‘one has a reason for something ... [if] it would advance some aim or purpose that one has’ then that is an instrumental reason for action.²⁷ On a substantive account of rationality ‘when we are aware of facts that give us certain reasons, we ought rationally to *respond* to these reasons’.²⁸ John Broome notes an objection to the ‘correct response’ thesis. He states:

On some occasion, there might be a reason for you to achieve something but, without any irrationality on your part, you might not believe this reason exists. If you do not believe it exists, then you might well not respond correctly to it, and your failure will not imply any failure of your rationality. Therefore, rationality cannot consist in responding correctly to reasons.²⁹

Some find this epistemological objection compelling and therefore adopt the view ‘that rationality consists in responding correctly to the reasons that you *believe* to exist’.³⁰ Thomas Scanlon, however, argues against this view using the example of

²⁶ Thomas Scanlon, *What We Owe Each Other* (Harvard University Press 2000) 19.

²⁷ *Ibid* 64.

²⁸ Derek Parfit, *On what matters*, vol.1 (Oxford University Press 2011) 111.

²⁹ Broome J, ‘Is Rationality Normative?’ 2007 *Disputatio* 2:23 167

³⁰ *Ibid* 167.

buying a new computer because I have a desire for a new machine in ‘the directed-attention sense’.³¹ He argues that such a ‘state [of desire] can occur ... even when my considered judgement is that I in fact have no reason to buy a new machine’.³² As desires can remain even we judge that we do not have a sufficient reason for the object of the desire Scanlon concludes that desires play no role in reason. Thus according to him:

[Desires] are instances of an agent’s identifying some other considerations as reasons, and derive their reason-giving force from a combination of these reasons and the agent’s decision to take them as grounds for action.³³

Thus Scanlon concludes that ‘a rational person takes herself to have a reason for an action ... without any further appeal to desire’.³⁴ I must confess that I find the idea of substantive reasons to which agents can respond correctly extremely implausible because reasons are contingent on our goals and preferences. The fact that it is raining outside is not a reason for me to wear waterproof clothing unless I want to avoid getting wet. If I enjoy standing in the rain and getting soaked then the fact that it is raining is a reason for me *not* to wear waterproofs. If I am indifferent to rain then the fact that it is raining is not a reason for me to do or not do anything. Reasons cannot exist independently from the goals of an agent – the reason-giving force of certain considerations is based upon what we want to achieve and how we go about it. It is for this reason that I am highly sceptical of the notion of substantive rationality.

Thus substantive rationality seems to require reasons that are independent of any particular agent that are objective features of the world. This idea would seem to open the door to claims that better decisions can be made for people both in terms of their actions and in terms of their goals. This would imply an imperialistic attitude towards rationality under which agents can be encouraged or forced to respond *correctly* to the right and relevant reasons. In any case, this thesis uses the methodology of practical rationality to determine better and worse outcomes – if there

³¹ Scanlon (n26) 43: Scanlon defines directed-attention desires as a ‘person’s attention is directed insistently toward considerations that present themselves as counting in favour of P’ 39.

³² Ibid 43.

³³ Ibid 45.

³⁴ Ibid 77.

are such things as substantive reasons (of which I am very doubtful) they are not relevant here.

In developing guidance and in finding practical solutions to interaction problems one *ought* to follow practical rationality. But why should we follow practical rationality? We might as well ask ‘why one should enter on the theoretical task of deliberation’ – the answer is that the ‘point of the theoretical is to give guidance in the practical’.³⁵ Consequently, we should follow practical rationality because this guides us along the best route for achieving our aims. For this reason I use ‘an instrumental conception of practical rationality, according to which a choice is rational if and only if relative to the agent’s beliefs it is the most effective means for achieving the agent’s goals’.³⁶ Hence, practical rationality determines an agent’s most effective means for maximising the benefit received from co-operation consistent with other agents acting likewise. Furthermore, ‘the theory of rational choice to derive moral principles from a *morally neutral* choice situation’ thus escaping the problem of moral foundationalism that it seeks to resolve.³⁷ This conception of rationality allows us to use means-end reasoning to propose courses of action to agents within specific contexts without assuming that moral claims (in)validate particular actions.

Practical rationality identifies the best reasons we have for taking a particular course of action to achieve our goals – once this course of action has been identified then we *ought*, in the normative sense of ought, to follow it because we have identified that which best promotes the most beneficial outcome. The metaphysical claims of cognitivist morality, by contrast, would determine (un)acceptable reasons without reference to agents or their goals – if an action is wrong then it is forbidden regardless of the practical outcomes it may generate. Moral error theory specifically targets this kind of moral claim and reduces moral claims to personal sentiment *thus practical rationality gains priority over morality*. Under moral systems an agent may say I think *x* is wrong and I shall not do it but another agent can simply reply I think *x* is right and I shall do it. Practical rationality would take these conflicting viewpoints as given and then seek to provide guidance on behaviour that is beneficial for both agents – generally, by permitting the broadest range of actions possible. The

³⁵ Kurt Baier, *The Moral Point of View* (Random House 1966) 160.

³⁶ Peter Vallentyne, ‘Gauthier’s Three Projects’ in Peter Vallentyne (ed), *Contractarianism and Rational Choice* (Cambridge University Press 1991) 5.

³⁷ *Ibid* 3.

distinction between rationality and morality is crucial for the theory of rational constraint to solve co-ordination problems.

There are two further reasons for maintaining this distinction. First, treating rationality and morality as the same might lead to the impression that this work seeks to identify mind-independent moral rules. Given the adoption of moral error theory this would invite confusion. Second, if morality must also be rational then the moral element is superfluous for it is rationality that determines what counts as moral. Thus morality would not be playing any part in justifying the rules generated because the rules will have been developed through rational assessment.³⁸ Claiming that rationality and morality are the same does allow one to use moral language and to portray the rules as moral rules that may confer a strength and persuasiveness to those rules than would otherwise be the case. But, if the rational and the moral are one and the same this will either confuse what the justification of the rules actually is or it will give greater force to the rules through moral language, which are unjustified due to the arguments of moral error theory.

But is it just semantics whether we use *moral* or *rational*, or does it reflect some deeper conceptual conflict? Could we not say that the action which we have the best reason to do is morally right thing to do? There is an ambiguity here because what counts as *the best reason* will depend on what we include as a *reason*. If best reason includes moral reasons then reasoning becomes circular, as morality is already present thus determining what counts as a reason. If moral claims are excluded then morality is simply rationality (as discussed above). For example, treating embryos as a life equal or unequal to an adult is a prior moral position that will determine what reasons are included in making a decision but this reason (counting embryos equally) is not a *best* or *better* reason – if one does not count embryos as morally relevant then it is not a reason at all. Thus *best reasons*, in the absence of substantive universal account of reasons, will be pre-determined by moral claims and so will merely be a moral judgement reached within a particular moral framework.

The idea of *best reasons* merely raises the problem discussed above regarding the concept of substantive rationality. The contingent nature of what counts as a *reason* is why the idea of simply following the *best reasons* – which requires a substantive notion of reasons – is so implausible. In other words, to even accept moral

³⁸ This is a charge that can levelled against Kurt Baier, *The Rational and the Moral Order* (Open Court Publishing 1995).

claims as reasons for or against particular actions an agent must already accept that moral position – thus what may count as the best reasons to one will not count as a reason at all to another. Consequently what view we hold of the relationship between rationality and morality is an issue of more than mere semantics, it is a fundamental issue concerning how we construct the world around us and how we determine what considerations we must take account of.

Practical reasoning does not suffer from this flaw because it only includes reasons that affect the outcome and success of our goal and these factors are determined by the physical circumstances in which actions take place. For example, the fact that it is raining outside might be a reason to take an umbrella for some but if I like the feel of rain against my face then it is a reason for me not to take an umbrella. Or the fact that an embryo is biologically human is irrelevant insofar as the success of gestation and the fulfilment of the parental goal are concerned. In practical rationality *reasons* are those and only those factors that are relevant to the production of certain outcomes – it is this that makes practical rationality more useful in setting out how to resolve inter-agent conflict but less useful in the determination of goals. Nevertheless, I see this as a strength of practical rationality because it explicitly acknowledges the role of goals in determining how we behave within a given context. Thus practical rationality can account for the moral element of agents because it is included in their goals – achieving those goals is the limiting condition for including certain factors as reasons. Additionally, by treating agents as prior to the system we ensure that it is agents who determine what constraints they are subject to.

Practical rationality places reason above sentiment morality (which is the only viable form of morality after the coup de grace of moral error theory) – two agents may disagree about whether action X is permissible or forbidden but neither view can be enforced on the basis of morality so they must reach a practical solution concerning how they regulate their behaviour towards each other. Finding this solution is, in a nutshell, the project that this thesis addresses.

Research Questions

This thesis begins by defining the role that moral claims can play as a methodology for conflict resolution amongst a heterogeneous population – this explains both the limits of morality and the need for an alternative system that avoids these problems

(chapter one). Next I present an outline of the alternative needed and explain how it avoids the problems of moral systems (chapter two) which leads to the application of this approach to human rights law (chapter three). Finally, I show why moral claims are ineffectual as they can be challenged on metaphysical grounds which prevent their use as conclusion statements or factors applicable to all agents (chapters four and five). The following are the research questions that this thesis seeks to answer:

- 1) Chapter 1 *Can Morality Justify Population-wide Constraint?*: This chapter explains why we cannot use moral claims to resolve inter-agent conflict and why moral error theory should be our start point. By setting out the limitations of moral systems this chapter sets the stage for the introduction and adoption of rational constraint theory.
- 2) Chapter 2 *Can Rational Constraint Secure Co-operation?*: The second chapter of this thesis sets out how a system of rational constraint would work, the problems a self-regulating approach faces and why second-order constraints are needed and provide a superior system to self-regulation. This chapter provides the crucial lynchpin between contractarianism, the rational choice methodology and law. By explaining the need for second-order constraint and explaining how legal rules are a form of second-order constraint, legal theory and philosophy are brought together in a practical method of analysis.
- 3) Chapter 3 *Is a Waiver of Human Rights a more Rational and Beneficial Theory of Rights?*: Here I apply the analytical method developed in the previous chapters to human rights law as it relates to reproduction. The purpose of this chapter is to demonstrate the usefulness of the practical methodology of rational constraint. In doing so it will answer the question of whether an alternative construction of human rights will grant agents more freedom than they currently possess without threatening the entire system of human rights. This will demonstrate the applicability of both the theoretical framework and the practical methodology to the assessment of legal rules.
- 4) Chapter 4 *Potential Persons and the Welfare of the (Potential) Child Test*: This chapter presents the arguments against the commonly held view that those created can be harmed by that action thus demonstrating the inability of cognitivist moral claims to provide a compelling reason for all agents to comply with that conception of reality. This further demonstrates the need for regulation to adopt a

moral error theory approach to the creation and enforcement of legal rules. Consequently, constraint cannot be justified on the basis of harm to potential future people because it does not confer any benefit and agents (as a whole) have no reason to constrain their actions. This chapter includes a subsection that considers further the fundamental mutually exclusive axiomatic basis of impersonal and person-affecting harm.

- 5) Chapter 5 *Applying the Actual/Potential Persons Distinction to Reproductive Torts*: This chapter argues that reproductive torts are viable between existing individuals but they should not be used by the child who exists as a result of the tort action and that to apply tort to potential persons violates the theoretical underpinnings of tort law and contradicts the purpose of tort law itself as a means of regulating voluntarily created legal relationships between agents. Thus the constraints enforced through the civil process of tort law can only hold between the medical professionals and the parents, not between the medical professionals and the child, because the role tort law plays within the legal system is conceptually restricted to legal relationships between existing agents. This chapter has been published and so is not structured in the way I would now have structured it – the chapter primarily focuses on the impossibility of calculating and applying damages due to the non-existence of those harmed. The chapter would work better if the focus was on the fundamental role of tort as a mechanism for mediating between two existing agents who have a legal relationship prior to the event which has given rise to litigation. By focusing on the role of tort law in its capacity of redress it would explain both why the lack of harm and the impossibility of damages is an essential part of the process and it would explain why those created by reproductive actions should be excluded from the litigation process.

Philosophical Approach

The philosophical approach taken here uses moral error theory to justify a contractarian theory of rational constraint to justify second-order constraint on interaction. Under this approach constraint is measured by how successfully it promotes the preferences of agents thus leading agents – through rational choice – to voluntarily comply with the system of constraint out of self-interest and minimising

enforcement costs. Looking at the available ethical theories the only tradition which does not rely on any moral assumptions is the contractarian theory (which sits within the broader social contract tradition) inherited from Thomas Hobbes and, most recently, expounded by David Gauthier.

Moral Traditions and Contractarian Theory

The social contract tradition stands in contrast to the deontological and consequentialist ethical theories. The well known consequentialist Utilitarian theory relies on the concept of utility that provides an interpersonal method of measuring various outcomes. A Utilitarian identifies the outcome which produces the greatest utility and agents are morally required to bring about that outcome. A deontological account seeks to determine a coherent systematic set of correct moral principles which will then guide an agent deciding on how to behave. Both require at least one moral master principle as a keystone. Social contract theory differs from both because it generates rules through a hypothetical agreement amongst pre-theoretical agents – by which I mean agents whose capacities and features are determined prior to application of the social contract.

In the contractarian rational choice methodology set out in this thesis the aim is to ensure the maximisation of preference fulfilment. One might be tempted to think that preference maximisation is no different from Utilitarian maximisation of utility – after all in both systems we are trying to maximise the value produced in the outcome reached. But there are some fundamental differences between the two systems. While contractarian practical rationality does seek to maximise the payoffs for each agent, maximisation plays an instrumental function in ensuring the mutual agreement of all participants in a system of constraint. In other words, maximisation is a method for reaching agreement, not the reason for agreement itself nor the purpose of this system. The reason for agreement is the mutual benefit that accrues to each agent – maximisation merely seeks to produce the highest individual benefits. The benefit that agents receive from participation is the reason for compliance and it is rational to maximise those benefits to the extent possible within a framework of other agents acting similarly. Interpersonal comparison plays a role in trying to identify relative concessions between individuals but the preferences of each agent are not simply the same value repeated across all agents and they cannot be substituted for each other.

Moreover, agents acting on the basis of their self-interest will not accept sacrificing themselves to benefit others simply to produce greater total utility because each agent's utility is their sole concern and because utility is not universal amongst all agents preventing the utility that accrues to one agent from being transferred to another. For example, my preference for reading novels cannot be substituted for someone else's preference for watching football because those preferences are our own – they do not represent some universal utility or value. In contrast, Utilitarianism and other consequentialist systems have as their sole purpose the maximisation of utility regardless of the effects on any particular individuals because the metric of utility is the same for all individuals. This treats utility as universal and the same for all agents permitting the transfer of utility between agents and its focus on total utility means that individuals do not matter. Contractarian rational constraint theory places individual agents centre stage and makes the fulfilment of their preferences a prerequisite for co-operation. The role of the individual as central to the determination of outcomes is the fundamental difference between contractarian rational constraint and consequentialist theories.

However, contract theory has developed along two distinct paths: contractualism is based on the assumption of inherent moral equality between agents while contractarianism assumes that agents are engaged in competition with each other. Contractarianism is to be preferred over contractualist, deontological and consequentialist theories because it alone derives a theory of constraint without *any* moral assumptions. Instead it relies on particular *empirical* and *descriptive* assumptions about agents and the context of interaction. These assumptions attempt to capture the salient features of the lives of human agents – primarily the limited power of an isolated individual and the correspondingly greater power of groups of coordinated agents. Divergence between contractarian theory and how actual agents behave can also inform our understanding of human psychology and constraint systems.

Competition is crucial to understanding the need for constraint. In *Leviathan*, Hobbes describes the situation of competition as the condition of 'every man, against every man'.³⁹ According to Hobbes there are 'three principal causes of quarrel ... [f]irst, competition; secondly, diffidence [distrust]; thirdly, glory. The first, maketh

³⁹ Thomas Hobbes, *Leviathan* (first publish 1651, Oxford University Press 1996) 84.

men invade for gain; the second, for safety; and the third, for reputation'.⁴⁰ However, people seek peace (co-operation) because of the 'desire of such things as are necessary to commodious living ... [a]nd reason suggesteth convenient articles of peace, upon which men may be drawn to agreement'.⁴¹ Hobbes proposed two laws of nature which would permit co-operation. The first, 'that every man, ought to endeavour peace, as far as he has hope of obtaining it; and when he cannot obtain it, that he may seek, and use, all helps, and advantages of war'.⁴² The second law supports the first for it requires that individuals 'lay down this right to all things; and be contented with so much liberty against other men, as he would allow other men against himself'.⁴³

This mutual laying down of the unlimited right in nature is performed through a covenant. A covenant is a hypothetical agreement between agents in which they agree to mutually constrain their conduct for the sake of peace and security – for Hobbes this meant empowering an all-powerful sovereign to enforce the covenant. In modern contractarian parlance a covenant is a social contract or agreement. The mutual acceptance of constraint is the fundamental basis of contractarian theory – 'each party to the covenant agrees not to oppose the exercise of some right by the other, and this is achieved by laying down his own corresponding right'.⁴⁴

David Gauthier's contractarian theory, set out in his seminal work *Morals by Agreement*⁴⁵, follows the Hobbesian approach in making no moral assumptions and does not presuppose a substantive account of an agent's values. He argues that through co-operation agents can offset the costs of competition *and* produce additional benefits that otherwise would not exist – facts about the human condition provide the 'basis or ground of the reason' for co-operation.⁴⁶ Gauthier calls the additional 'gains which co-operation may bring ... the co-operative *surplus*'.⁴⁷ Co-operation is beneficial because the 'sources of satisfaction and dissatisfaction are not in fixed supply, so that by appropriate interaction overall costs may be lessened, and

⁴⁰ Ibid 83.

⁴¹ Ibid 86.

⁴² Ibid 87.

⁴³ Ibid 87.

⁴⁴ David Gauthier, *The Logic of Leviathan* (Clarendon Press 2011) 41.

⁴⁵ (Clarendon Press 2001) (n11).

⁴⁶ Baier, *The Rational and the Moral Order* (n38) 34.

⁴⁷ Gauthier, *Morals by Agreement* (n11) 130.

overall benefits increased'.⁴⁸ Self-interest in receiving benefits drives agents to the table and ensure compliance with the agreed upon constraint.

However, Gauthier differs significantly from Hobbes in arguing that it is not fear that drives agents to form an agreement. He seeks to provide an internal source of compliance through rational choice theory rather than by the threat of a Hobbesian sovereign.⁴⁹ Hobbes argues that 'nothing is more easily broken than a man's word' and consequently 'bonds ... have their strength ... from fear of some evil consequence upon rupture'.⁵⁰ Nevertheless, Hobbes does not intend a totalitarian regime. Rather, Gauthier suggests that Hobbes intends an 'authoritarian but benevolent' sovereign because the sovereign only does what is necessary for the maintenance of co-operation and no more.⁵¹ In contrast, Gauthier seeks to do the following: defend 'a specific contractarian theory of morality; [defend] ... a specific instrumental theory of rationality; and [defend] ... the claim that under a broad range of circumstances, rationality requires one to act morally'.⁵² This differs from Hobbes by relying on self-constraint on the part of participating agents rather than relying on the fear induced by the external threat of a sovereign's punishment. Unfortunately, Gauthier fails to provide a self-regulating form of constraint but this failure explains the need for second-order constraint to make predatory strategies irrational and bring about a more optimal outcome.

This contrasts with the Rousseauian *contractualist* tradition, of which John Rawls and Thomas Scanlon are the most prominent theorists. Jean-Jacques Rousseau founds his social contract theory on the claim that agents 'all being free and equal' must agree to constraint.⁵³ The conception of agents as equal is the foundational moral claim upon which the whole contractualist methodology rests – for example, Scanlon holds that all agents are 'rational creatures who recognise many of the same reasons

⁴⁸ Gauthier, *Morals by Agreement* (n11) 114.

⁴⁹ Note that while Hobbes identifies a sovereign as either one man or an assembly he favours a monarchical sovereign of one man over an assembly because an assembly leads to inconsistency whereas the monarch does not (so Hobbes claims): see Hobbes (n41) 124-5.

⁵⁰ Hobbes (n41) 88.

⁵¹ Gauthier, *The Logic of Leviathan* (n46) 139. In part IV: 'Theory of Authorization' of *The Logic of Leviathan* Gauthier attempts to provide an account of the sovereign as constrained by their purpose to ensure peace and for those with an interest in exploring this neo-Hobbesian theory of sovereignty this section to be of great assistance.

⁵² Vallentyne, 'Gauthier's Three Projects' (n36) 11.

⁵³ Jean-Jacques Rousseau, *The Social Contract* (first published 1762, Oxford University Press 2008) 46.

and can recognise the value of each other'.⁵⁴ John Rawls exacerbates the moral dependency of contractualism by making the following assumptions: agents 'are not presumed to be egoistic or selfish'⁵⁵, a 'rational individual does not suffer from envy'⁵⁶, 'the parties do not seek to confer benefits or to impose injuries on one another'⁵⁷ and 'their capacity for a sense of justice insures that the principles chosen will be respected'.⁵⁸ All three – Rousseau, Scanlon and Rawls – rely on an element of moral equality between agents that holds *prior* to the development of constraint thus relying on a foundationalist conception of morality precluded by the adoption of moral error theory. Furthermore, 'Rawls and Scanlon ... have regarded themselves as moral objectivists ... and have linked their theorising in some fashion to Kant'.⁵⁹ Consequently, the dependence of these theories on the fundamental objective moral claim to equality rules them out as the starting point for a system of constraints because it contradicts moral error theory and it requires that agents accept equality before constraint can be developed.

The contractarian tradition *can* function as a starting point because it relies on a descriptive premise of human agents. Hobbes states that '[n]ature hath made men so equal' through the threat that each poses to each other and this explains why a lone agent is better off engaging in co-operation.⁶⁰ This is not equality in any moral sense, rather it is more a *factual equality* given the limitations of human beings. Moreover, it places self-interest at the forefront of rational choice. Contrary to the notion of morality as restricting self-interested action, contractarianism makes self-interest the central method by which agents identify the best action for themselves including compliance with constraint. After all, 'it is clear that it is in the interest of everyone alike if everyone alike should be allowed to pursue his own interests provided this does not adversely affect someone else's interests'.⁶¹ Yet, agents interact with other self-interested agents so they must each take account of the other because the actions of one agent in strategic interaction affect the actions of other agents. Actions that

⁵⁴ Scanlon (n26) 77.

⁵⁵ John Rawls, *A Theory of Justice* (Harvard University Press 1999) 111.

⁵⁶ *Ibid* 124.

⁵⁷ *Ibid* 125.

⁵⁸ *Ibid* 125.

⁵⁹ Jean Hampton, 'Two Faces of Contractarian Thought' in Peter Vallentyne (ed) *Contractarianism and Rational Choice* (Cambridge University Press 1991) 32.

⁶⁰ Hobbes (n41) 82: It is debatable whether this is equality as we commonly understand it, for we normally consider that people are equal regardless of their relative strength to each other not because of it.

⁶¹ Baier, *The Moral Point of View* (n35) 107.

encourage defection from co-operation will undermine the system of constraint which confers benefits on all agents.

Consequently, prior to the social contract ‘notions of right and wrong, justice and injustice have there no place’.⁶² Right and wrong, justice and injustice, notions of permitted, forbidden and obligation are all absent because these concepts are a consequence of the social contract, *they are not the foundation for it*. This theme is also present in the contractualist strand of social contract theory. Rousseau suggest that ‘the social order is a sacred right, and provides a foundation for all other rights’ yet it ‘is a right that does not come from nature; therefore it is based on agreed convention’.⁶³ Rawls states that individuals should ‘recognise certain rules of conduct as binding’ which ‘specify a system of co-operation designed to advance the good of those taking part in it’.⁶⁴ Scanlon has a slightly different conception – he suggests that the ‘parties whose agreement is in question are assumed not merely to be seeking some kind of advantage but also to be moved by the aim of finding principles that others, similarly motivated, could not reasonably reject’.⁶⁵ The accrual of mutual benefit to each agent provides the motivation for forming and complying with a social contract. Yet this contradicts the principle of equality that contractualists espouse because the equality of each agent is not relevant to receiving benefit – either another agent can confer a benefit or not, if they cannot then equality is beside the point.

The contractarian approach seems contrary to all notions of morality for we ‘ordinarily suppose that someone who acts solely out of self-interest does not act from a moral reason’.⁶⁶ This further highlights the distinction between rationality and morality – although some still take rationality and morality *to be one and the same*. Gauthier can be criticised for treating them as synonymous. He states that ‘the rational principles for making choices, or decisions among possible actions, include some that constrain the actor pursuing his own interest in an impartial way. These we identify as moral principles’.⁶⁷ However, for the reasons discussed earlier I treat rationality and morality as strictly separate concepts.

Exclusivity of the Social Contract

⁶² Hobbes (n41) 85.

⁶³ Rousseau (n55) 46.

⁶⁴ Rawls (n57) 4.

⁶⁵ Scanlon (n26) 5.

⁶⁶ Gilbert Harman, *The Nature of Morality* (Oxford University Press 1977) 151.

⁶⁷ Gauthier, *Moral by Agreement* (n11) 3.

By now it should be clear that only existing agents have a primary position in contractarian theory – only they can constrain their actions and only they can make their participation in mutually beneficial co-operation conditional on another’s self-constraint. Constraint and co-operation ‘is possible only among contemporaries who actually interact’, thus we reach the main criticism of the social contract tradition.⁶⁸ Contractarian theory includes only a very limited class of beings within the system of constraint – those who are existing, atomistic, rational, autonomous beings. This is the ‘radical individualism’ of Hobbesian theory, although it afflicts all social contract theories because they all use the notion of mutuality to justify constraint.⁶⁹ Clearly ‘[a]nimals, the unborn, the congenitally handicapped and defective, fall beyond the pale of a morality [or rational constraint] tied to mutuality’.⁷⁰ Furthermore, while mutuality provides a rational foundation of constraint ‘it gives us less, in providing that foundation only within the confines of mutual benefit’.⁷¹ Gauthier and Rawls attempt to overcome this problem by extending the coverage of constraint to include those who fail to meet these requirements.

Gauthier claims that ‘[m]utually beneficial co-operation directly involves persons of different but overlapping generations, but this creates indirect co-operative links extending throughout history’.⁷² Yet as Gauthier notes, this claim is valid only if ‘we suppose that moral relationships among persons of different generations require[s] an affective basis’.⁷³ Two interpretations of this are possible. First, our preferences towards future people *result* in interaction between present and future people. Second, our preferences towards future people give *us* reason to act to bring about certain outcomes as though they are participants.

Adopting the first interpretation for constraint completely frustrates the contractarian programme of generating constraint through mutually beneficial interaction with other similar agents – intergenerational relationship beyond co-temporal generations are entirely one way. It also seems to be ad hoc – if an affective basis is suitable for future generations then why should it not form the basis of

⁶⁸ Gauthier, *Morals by Agreement* (n11) 298.

⁶⁹ Hampton (n61) 46.

⁷⁰ Gauthier, *Morals by Agreement* (n11) 268: it was tempting to omit ‘and defective’ as this language outdated and derogatory but it has been left in so that it is not seen as an attempt to sanitise Gauthier’s statement.

⁷¹ *Ibid* 268.

⁷² *Ibid* 299.

⁷³ *Ibid* 298.

constraint itself? Moreover, '[e]ven if, from time to time, there might be a powerful appeal made to sympathy, the fact that at other times there might not be enough' renders it unsuitable for justifying constraint.⁷⁴ Gauthier opens up the possibility of constraint founded on affective attitudes but only so that future generations are included. This cannot be done without also raising the possibility of constraint being entirely justified on an affective basis or else undermining the basis of an affective claim.

Rawls takes a different approach. He states that 'moral personality refer[s] to a capacity and not the realisation of it' thus a 'being that has this capacity, whether or not it is yet developed, is to receive the full protection of the principles of justice'.⁷⁵ But he makes clear that 'this interpretation of the requisite conditions [for inclusion] seems necessary to match our considered judgements'.⁷⁶ Furthermore, this contradicts both the requirement that '[t]hose who can give justice are owed justice'⁷⁷ and that 'the principles of justice [are] ... those which rational persons concerned to advance their interests would consent to as equals'.⁷⁸ The requirement of mutuality cannot be reconciled with the view that potentiality (to use Rawls's term) is sufficient for inclusion in justice. Even though Rawls is theorising about a decision made by agents who have no specific information about themselves – including which generation they are born into – both Rawlsian and actual agents have no reason to consider non-agents as participants. Indeed, they have reason to exclude them because they cannot *give justice* and this includes future generation with whom one cannot interact. As Peter Singer puts it:

Rawls deals with infants and children by including *potential* moral persons along with actual ones within the scope of the principles of justice. This is an ad hoc device, confessedly designed to square his theory with our ordinary moral intuitions.⁷⁹

⁷⁴ Brassington (n13) 95.

⁷⁵ Rawls (n57) 445-6.

⁷⁶ Ibid 446.

⁷⁷ Ibid 446.

⁷⁸ Ibid 17.

⁷⁹ Peter, Singer, *Practical Ethics* (3rd edn, Cambridge University Press 2011) 18-9.

Thus the extension of social contract theory to non-agents cannot be seen as anything more than an *ad hoc* move to accommodate widespread intuitions regarding the status of infants, severely mentally handicapped adults and potential persons.

Yet this leaves non-agents in a precarious situation as they can be included only to the extent that they are a component of an agent's preferences. Christopher Morris criticises contract theory on this basis for taking 'the metaphor of a social contract too literally'⁸⁰ because mutuality requires that 'only agents who were members to the contract would have moral standing'.⁸¹ He proposes a solution on the basis of *secondary moral standing*. A and B are agents and C is the offspring of B, B therefore makes participation in the constraint system conditional on C also being protected by those constraints. Given the moral theoretic perspective adopted by this thesis we should eliminate the word moral in this concept. Notwithstanding this terminological liquidation, if the protection of non-persons is included in an agent's preferences then it is rational for that agent to require the constraint system to consider the non-person as well – that is, an agent will make their participation contingent on some protection for the non-agent (i.e. parents will only participate if their children are protected by the system of constraint).

This solution is not so easily applied in relation to future persons whom we cannot be sure we affect but it would justify permitting individual freedom in this area because of those agents preferences regarding offspring – although this would be limited to *their* offspring, not the offspring of other agents. By this mechanism contractarian theory would extend constraint beyond mutuality because non-agents – whether existing non-persons, future people, animals or even environmental conditions – would be included albeit only to the extent that agents have preferences about them.

However, the 'exclusion of moral and tuistic motivations is a deep feature of the contractarian strategy for defeating moral scepticism ... [because it] would obviously beg the question to take into account moral preferences in the argument'.⁸² On this basis Morris's notion of secondary-standing would violate this restriction by allowing a concern of others to be included in an agent's preference set. But this

⁸⁰ Christopher Morris 'Moral Standing and Rational-Choice Contractarianism' Projects' in Peter Vallentyne (ed), *Contractarianism and Rational Choice* (Cambridge University Press 1991) 83-85, 83.

⁸¹ Ibid 85.

⁸² David Copp, 'Contractarianism and Moral Skepticism' in Peter Vallentyne (ed) *Contractarianism and Rational Choice* (Cambridge University Press 1991) 222.

exclusivity claim overstates the case. For moral and tuistic motivations are insufficient to ground a system of constraint due to the limited extent of our sympathies – thus the claim regarding ‘moral’ preference is not as strong as it seems. Moreover, this would only be a problem if we took moral preferences to be cognitivist mind-independent claims. If moral preferences are merely the expression of an attitude or sentiment then they must be included because they will affect an agent’s practical reason. Nevertheless, this does not entail assuming that all agents are moral – *even less that they all have the same moral preferences.*

For some this will not go far enough because it only assigns worth to non-agents derivatively. For others, this will be too much because it extends the system of rational constraint to *whatever* agents consider valuable. But if we are to consider a contractarian model that makes no moral assumptions, that attempts to use a descriptive account of agents, that treats their preferences as given and that is concerned with how agents will practically interact with each other, then what agents assign value to becomes an *integral* part of the system of constraint. If their preference concerning a non-agent is sufficiently strong then they will make the treatment of that non-agent a condition for their participation – and only the specific preferences of an agent will tell us if that is the case. While most may consider future people to be relevant in determining our actions, many will not care at all whether future people exist or in what condition. Others may seek to prevent the existence of future people entirely – the Voluntary Human Extinction Movement springs to mind.⁸³

Furthermore, while agents can demand that all of their preferences form the basis for the system of generally applicable constraint this outcome will not be realised because it amounts to making one agent the sole arbiter of constraint which gives all other agents a reason to oppose that system. Consider the following: anyone who required that all of their preferences were enforced through the constraint system would be unable to interact with others without the extensive use of force which would give others a reason to exclude such an agent from co-operation. Thus while preferences could include any value this would not *necessarily* justify constraint on others in interaction, although it would give the agent themselves a reason for self-constraint and may require modification of the constraints.

⁸³ Voluntary Human Extinction Movement <http://www.vhemt.org/>

For these reasons using the notion of *considered preferences* allows preferences to act as a place-holder for *whatever* preferences – whether beliefs or desires – agents possess. Considered preferences can include such a variety of preferences because the only requirements it imposes are completeness, transitivity, monotonicity and continuity⁸⁴. Furthermore, social contract theories treat ‘population size [and make up] as given and fixed’ – making this theory extremely useful because we can examine particular agents in particular contexts.⁸⁵ Consequently, we can then examine how such actual agents would function in interaction through practical rationality and propose what constraints would be accepted or what constraints should be modified or rescinded in the actual socio-political environment in which we live. This is of particular use when analysing legal rules in specific jurisdictions that are nestled within a broader socio-political context.

Legal Approach

Tracing legal theory

The philosophical approach outlined above leads to a unique conception of legal rules as second-order constraints on interaction to bring about the most optimal co-operative outcome based on empirical data concerning those to be subjected to legal regulation. As a result of this practical focus engendered by the contractarian theory my legal approach can be classified as a version of Legal Realism. Legal Realism rejects the role of metaphysical concepts such as a justice in decision making by judicial officials and focuses on the actual decisions they make and the impact those decisions have for legal subjects. Thus Legal Realism treats law as a social phenomenon shorn of metaphysical mysticism and analysed on the empirical outcomes it produces through the decisions of judicial officials. This is in line with the contractarian approach and the methodology of rational choice used which analyses the enforced laws and requires that legal rules bring about a more optimal outcome than the alternatives of no regulation or different regulatory forms. The integration of contractarian philosophy, the methodology Legal Realism and rational choice produce the theory of rational constraint. When applied to legal rules rational constraint treats law as second-order constraints on interaction which is justified by

⁸⁴ Gauthier, *Morals by Agreement* (n11) 39-40.

⁸⁵ David Heyd, *Genethics Moral Issues in the Creation of People* (University of California Press 1992) 42.

the benefits agents receive from compliance and which is judged by the actual effects of those laws.

The increasing complexity of legal systems, a consequence of the populations they regulate becoming more diverse, has led to a more distributed system of enforcement – no longer is the sovereign the arbiter of justice and law. Instead agents are encouraged to comply with regulation because this is to their benefit and because enforcement is applied more consistently and frequently. This mirrors the development from Hobbesian political philosophy concerning the role of the Sovereign as the means of enforcing constraints upon subjects⁸⁶ to Gauthier's suggested internal self-regulation of constraint. Sovereigns and their nations – both legally and politically – have been transformed from autocratic regimes where the authority of power is the absolute power of command to some form of oligarchic constitutional system, whether that is a democratic state along the lines of 'Western' nations or the authoritarian systems of some states.⁸⁷ This transformation has been accompanied by a changing theoretical discourse surrounding political and legal systems, thus the Hobbes-Gauthier model of contractarianism has parallels within legal theory. Tracing the development of legal theory from John Austin through to Herbert Hart, Ronald Dworkin and Lon Fuller will bring out these similarities and situate the interpretation of law that this thesis uses.

Austin's Command Theory of Law

Austin proposed a theory of law that bears a striking resemblance to Hobbes's political theory. Austin's basic conception of law is that 'Laws proper, or properly so called, are commands'⁸⁸ and 'being a *command*, every law properly so called flows from a *determinate* source'.⁸⁹ The 'author from whom it proceeds is a *determinate* rational being, or a *determinate* body or aggregate of rational beings'.⁹⁰ For Austin this is a sovereign because 'a monarch or sovereign number in the character of a political superior'.⁹¹ A sovereign can command but cannot be commanded making them the highest authority for issuing commands. Enforcement of these commands is achieved by the use of sanctions or, as Austin puts it, the commanded 'is obnoxious to

⁸⁶ Hobbes (n41) 115-122.

⁸⁷ Belarus, Saudi Arabia, China, Russia, Iran and many others come readily to mind.

⁸⁸ John Austin, *The Province of Jurisprudence Determined* (Hackett Publishing Company 1998) 1.

⁸⁹ *Ibid* 133.

⁹⁰ *Ibid* 133.

⁹¹ *Ibid* 134.

an evil which the former [the issuer of the command] intends to inflict in case the wish be disregarded'.⁹² Both Hobbes and Austin view a sovereign as the necessary and sufficient means of securing obedience and peace through the use of force and fear of evil consequence that the sovereign *alone* can authoritatively justify. According to Hobbes '[t]he legislator in all commonwealths, is only the sovereign' and 'the legislator, is he that maketh the law'.⁹³ Most importantly, '[t]he sovereign of a commonwealth, be it an assembly, or one man, is not subject to the civil laws'.⁹⁴

Austin's command theory of law has been criticised for its simplicity for it does not allow us to distinguish between the commands of a sovereign and a gunman, nor does it account for laws which are *not* backed by threat.⁹⁵ Hart points out that 'there are other varieties of law, notably those conferring legal powers ... which cannot, without absurdity, be construed as orders backed by threats'.⁹⁶ Furthermore, 'the law's strictures – and its sanctions – are different in that they are obligatory in a way that the outlaw's commands are not. Austin's analysis has no place for any such distinction'.⁹⁷ This leads Hart to distinguish between 'the assertion that someone *was obliged* to do something and the assertion that he *had an obligation* to do it'.⁹⁸ The former is 'a psychological one referring to the beliefs and motives with which an action was done' – such as being motivated by the gunman's threat of violence.⁹⁹ To say 'a person had an obligation ... remains true even if he believed (reasonably or unreasonably) that he would never be found out and had nothing to fear from disobedience'.¹⁰⁰ This distinction rests on, what Dworkin calls, a rule 'being *normative*, by setting a standard of behaviour that has call on its subject beyond the threat that may enforce it'.¹⁰¹

The critique of Austin's command theory boils down to this: Austin sees legal rules as only being obeyed because of the threat of a sanction, while Hart and Dworkin see legal rules are being obeyed because they have some *normative* claim on agents. The fundamental difference between Hobbes and Austin on the one hand and

⁹² Ibid 133.

⁹³ Hobbes (41) 176.

⁹⁴ Ibid 176.

⁹⁵ Herbert Hart, *The Concept of Law* (Oxford University Press 2012) 19 and Raymond Wacks, *Understanding Jurisprudence* (Oxford University Press 2012) 73-4.

⁹⁶ Hart 79.

⁹⁷ Ronald Dworkin, *Taking Rights Seriously* (Bloomsbury Academic 2013) 34.

⁹⁸ Hart (n97) 82.

⁹⁹ Ibid 82.

¹⁰⁰ Ibid 83.

¹⁰¹ Dworkin (n99) 35.

Gauthier, Hart and Dworkin on the other is that both Hobbes and Austin see compliance with constraint as achievable only through the threat of coercive force, while for Gauthier, Hart and Dworkin compliance is achievable through the normative justification of constraint. The key discussion for the latter theorists is *justifying* constraints such that even in the absence of an enforcement mechanism agents should and will comply with them.

Hart and Dworkin on Law – A Comparison with Rational Constraint

Hart proposes the notion of a *rule of recognition* which confers normativity to legal rules, which in turn rest upon the distinction between primary and secondary rules. Primary rules are those rules under which agents ‘are required to do or abstain from certain actions’.¹⁰² Secondary rules are those rules under which agents ‘by doing or saying certain things introduce new rules of the primary type, extinguish or modify old ones, or in various ways determine their incidence or control their operations.’¹⁰³ For example, the rules for adoption are primary rules while the mechanism by which the rules of adoption change is the secondary rule. The notion of primary and secondary rules has ‘central place [in jurisprudence] because of their explanatory power in elucidating the conceptions that constitute the framework of legal thought.’¹⁰⁴

The crucial explanatory power of Hart’s primary and secondary rule thesis comes to the fore when he introduces the secondary rule of recognition which provides the ‘authoritative criteria for identifying primary rules of obligation.’¹⁰⁵ The rule of recognition – which ‘may take any of a huge variety of forms, simple or complex’¹⁰⁶ – is the most important rule because ‘the rules [of a legal system] are now not just a discrete unconnected set but are, in a simple way, unified’.¹⁰⁷ Hart’s rule of recognition justifies the primary rules that result from it. A contractarian theory of rational constraint can – in principle – be used to justify or critique *any* rule. Rational constraint would be able to bypass a rule of recognition – after all, on Hart’s positivist account the rule of recognition is the rule that *happens* to govern primary rules rather than being the best, most effective or otherwise determined optimal rule. On this ground alone Hart’s claim that the rule of recognition confers normativity can be

¹⁰² Hart (n97) 81.

¹⁰³ Ibid 81.

¹⁰⁴ Ibid 81.

¹⁰⁵ Ibid 100.

¹⁰⁶ Ibid 94.

¹⁰⁷ Ibid 95.

challenged – in the UK the monarch’s assent is required for legislation to be enacted but it would seem grossly medieval to say that the assent of a single person (who holds the office by happenstance) has the power to create normative laws. It may have no normative justification but may merely be the result of historic political circumstances. Yet, Hart relies on it to justify the obligations that legal rules impose upon legal subjects. Rational constraint on the other does not need a ‘rule of recognition’ to justify constraint because the validity of legal rules is directly tied to their effects upon legal subjects and the fulfilment of their preferences. Thus legal subjects play a direct active role in determining whether laws are justified or not rather than using an historical socio-political process to justify legal regulation.

In contrast to both Hart’s and my approaches Dworkin takes moral considerations to be prior to the formation of law. Dworkin first states legal positivism as follows: ‘Legal positivism assumes that law is made by explicit social practice or institutional decision’¹⁰⁸ and it ‘rejects the idea that legal rights can pre-exist any form of legislation’.¹⁰⁹ Thus ‘the rule does not [and cannot] exist before the case is decided [or before legislation is enacted]; [*but*] the court cites principles as its justification for adopting and applying a new law’.¹¹⁰ For Dworkin principles play a vital role in determining the law particularly as exercised in adjudication by judges. A principle is ‘a standard that is to be observed ... because it is a requirement of justice or fairness or some other dimension of morality’.¹¹¹ Principles, therefore, render legal positivism inadequate because legal positivism ignores the role principles play in the formation of legal rules. Consequently, it ‘forces us to miss the important roles of these standards that are not rules’.¹¹² This is particularly problematic because judicial systems regularly refer to principles such as the principle ‘that no man may profit by his own wrong’.¹¹³ Dworkin claims that the legal positivist account cannot explain the normativity of law by reference to *social rules* (i.e. the socially determined rule of recognition) because it misses the role of *principles*. The legal positivist ‘must concede that there are some assertions of a normative rule that cannot be explained as an appeal to a social rule’.¹¹⁴ Hart suggests this approach means that ‘for Dworkin the

¹⁰⁸ Dworkin (n99) 4.

¹⁰⁹ Ibid 6.

¹¹⁰ Ibid 45. My italics.

¹¹¹ Ibid 39.

¹¹² Ibid 38.

¹¹³ Ibid 39.

¹¹⁴ Ibid 72.

truth of any proposition of law ultimately depends on the truth of a moral judgement as to what best justifies' the rule.¹¹⁵

Dworkin's methodology identifies the principles to which law must conform and only then considers whether legal rules comply with those principles. He can, therefore, discuss the validity of laws without reference to the rule of recognition – he could, for example, critique the laws of the Russian Federation relating to the promotion of homosexuality on the grounds that this does not reflect a principle of equal treatment, fairness, justice or some other principle of morality. I mention this merely to point out that this is exactly the situation that the theory of rational constraint should resolve – for while we can disagree over what fairness or justice is we should, in principle, be able to identify the most rational judicial decision or legal rule.

Hart's positivist theory focuses on the practices of law and rules that govern the validity of legal rules while Dworkin argues that legal rules must be consonant with moral principles. The positivist tradition within which Hart works that limits the application of his theory to actual legal practice. This is not a failing of that theory but it does mean that Hart cannot say much about the extra-judicial principles that are articulated within legal discourse. He can only provide descriptive content to legal systems. However, Dworkin explicitly ties moral principles into the justification for legal rules, which is problematic for the reasons set out earlier concerning moral error. The rational constraint methodology of this thesis can provide extra-judicial justificatory analysis beyond the mere application of legal rules without resorting to moral principles which are unable to resolve or identify the most optimal rules. What Dworkin provides – which Hart cannot – is a normative grounding for legal rules. Yet the reliance on moral concepts weakens the applicability of his account. The theory of rational constraint *can* provide a non-moral normative grounding for legal rules because the normativity of legal rules derives from the mutual benefit attained by co-operating agents and is, therefore, a superior theoretical framework from which to evaluate legal rules.

Using the theory of rational constraint and treating law as a subdivision of constraint one can approach legal rules from the perspective of the legal subject and analyse whether the rule is one they would rationally agree to and under what

¹¹⁵ Hart (n97) 253.

conditions. This treats legal rules as second-order rational constraints based upon a rational choice methodology. Rational constraint theory asks whether the rule is mutually beneficial to all participating agents, whether there are alternatives that increase the benefits received and whether acceptance of the rule is conditional on other rules or benefits. Criminal law is not beneficial for those classed as criminals for it stops some agents from fulfilling their preferences – rapists presumably have a preference for that kind of sexual activity yet fulfilment of their preference in this regard is completely prohibited by the criminal justice system because other agents will rationally make freedom and protection from rape a condition of their participation. Thus a critical majority of agents will insist that a criminal justice system is adopted and enforced as a condition for their participation because criminal justice systems are beneficial for almost all agents and may be necessary for securing co-operation. I will now consider some of the interconnections between Hart's and Fuller's legal theories and the theory of rational constraint which will help clarify the empirical-descriptive basis of the rational choice methodology used here.

Links between Legal Theory and Contractarianism

The closest conjunction occurs with Hart's 'minimum content of natural law'¹¹⁶, which is an 'empirical version of natural law ... based on Hobbes'.¹¹⁷ This comprises the following: human vulnerability¹¹⁸, approximate equality¹¹⁹, limited altruism¹²⁰, limited resources¹²¹, limited understanding and strength of will.¹²² On this account law compensates for human frailty through a system that enforces 'conventional rules as a guide to their [agents] behaviour or standards of criticism'.¹²³ An agent can act in the knowledge that nearly all other agents will behave in certain predictable ways, while those who defect from the system will be punished. While these Hobbesian-derived conditions of human living makes an appearance in Hart's theory it does not seem to have great prominence which makes it unclear whether Hart thought these

¹¹⁶ Ibid 194.

¹¹⁷ Ibid 303.

¹¹⁸ Ibid 194.

¹¹⁹ Note that approximate equality is an empirical claim not a moral statement. As Hart puts it, 'no individual is so much more powerful than others, that he is able, without co-operation, to dominate or subdue them for more than a short period' Ibid 195.

¹²⁰ Ibid 196.

¹²¹ Ibid 196.

¹²² Ibid 197.

¹²³ Ibid 257.

conditions *justify* legal systems or are simply an a socio-historical explanation of them. In contrast, they play a vital role in rational constraint theory because these features of human existence explain the need for constraint and the need for sufficiently widespread compliance to be effective at minimal cost in terms of enforcement mechanisms. Thus legal rules are justified by how efficiently they compensate agents by conferring benefit for the restrictions imposed upon them – regardless of other factors such as a rule of recognition, principles of morality or even the particular form of governance in use.

Fuller provides a theory of law closer to Gauthier’s contractarian theory because they both treat morality and rationality as entwined – that is, what is rational is moral and what is moral is rational. Fuller proposes two conceptions of morality to underpin the law – the morality of aspiration and the morality of duty. Morality of aspiration concerns ‘the morality of the Good Life, of excellence, of the fullest realisation of human powers’, whilst the morality of duty concerns ‘the basic rules without which an ordered society is impossible’.¹²⁴ While the conflation of morality and rationality is not mere semantics, as I have suggested, it is unhelpful because morality plays no significant role in determining what is rational if they are one and the same. If, on the other hand, we say that rationality must conform to morality then they *are* distinct separate concepts but that reintroduces the problems that motivated the move to the methodology of practical rational choice in the first place (namely the fact that moral claims are not truth-apt).

Fuller correlates these two moralities with the law and constructs a moral scale of classification. This scale ‘begins at the bottom with the most obvious demands of social living and extends upwards to the highest of human aspiration’.¹²⁵ Fuller claims that ‘[i]t is obvious that duties, both moral and legal, can arise out of an exchange’ and the ‘affinity between duty and exchange’ is found ‘in the relationship of reciprocity’.¹²⁶ Reciprocity provides the foundation for exchange, exchange provides the foundation for duty and duty provides the foundation for constraint. Reciprocity for Fuller plays the same role as mutual benefits does for Gauthier. Reciprocity is equivalent to mutual benefit because both require that all participating agents gain from co-operation by conferring some benefits on each other. Assuming that agents

¹²⁴ Lon Fuller, *The Morality of Law* (Yale University 1964) 5.

¹²⁵ *Ibid* 10.

¹²⁶ *Ibid* 19.

will participate voluntarily only when they can benefit from interaction, ‘the relationship of reciprocity out of which the duty arises must result from a voluntary agreement’.¹²⁷

But agents ‘must in some sense be equal in value’.¹²⁸ This phrasing is sufficiently vague that it could mean equality in the Rousseauian sense of moral equality, the Scanlonian sense of mutual recognition or the Hobbesian sense of vulnerability. If the Rousseauian or Scanlonian senses of equality are adopted Fuller’s theory will fall victim to moral foundationalism – only the Hobbesian descriptive sense of equality (that is human frailty and limitation) avoids this. These requirements also entail that ‘the relationship of duty must in theory and in practice be reversible’, which means each agent must be able to perform any role required of any other agent.¹²⁹ Fuller’s account highlights the omni-directional role of reciprocity and exchange as the basis of self-constraint.

Fuller further posits that the society which best realises reciprocity, exchange and duty is ‘a society of economic traders’ whose members ‘enter direct and voluntary relationships of exchange’.¹³⁰ Gauthier also uses notion of ‘the perfect market ... [which] guarantees the coincidence of equilibrium and optimality’ as the ideal model for voluntary mutually beneficial constraint.¹³¹ Both Fuller and Gauthier use a hypothetical idealised economic state as the model for perfect constraint because of the roles played by voluntary agreement and disinterestedness between agents under an economic model. An economic agent does not act out of charity or moral concern and this is why mutual benefit is the foundation of constraint.

Richard Posner’s economic theory of law develops this approach to the fullest extent.¹³² Under Posner’s theory the law operates on the basis of wealth-maximisation because wealth is the medium through which agents express their preferences. Yet, these theories ‘are not value free, but are intimately related to and dependent upon modern American conservative ideology’.¹³³ Fuller and Gauthier reproduce the concepts of capitalist discourse by taking the economic methodology as *the* standard

¹²⁷ Ibid 23.

¹²⁸ Ibid 23.

¹²⁹ Ibid 23.

¹³⁰ Ibid 24.

¹³¹ Gauthier, *Morals by Agreement* (n11) 83.

¹³² See: Richard Posner, *Economic Analysis of Law* (Little, Brown & Company 1973) and Richard Posner, *Economics of Justice* (Harvard University Press 1998).

¹³³ Jack Balkin, ‘Too Good To Be True: The Positive Economic Theory of Law’ (1987) *Faculty Scholarship Series Paper* 290, 1447.

to which we should aim – it ignores the non-economic contexts and networks that agents are embedded within. We must be cautious of the economic model of constraint. Instead we should move towards a more complex realistic description of agents and interaction which can only be an improvement in determining practical solutions to co-ordination problems (despite the need for some idealisation to distinguish society as it *is* from society as it *could be*). Such a move would expand the usefulness of rational choice analyses in designing legal regimes and in guiding judicial practice.

Towards a political and legal theory of constraint

A contractarian theory of rational constraint can justify limitations on an agent's actions at the constitutional level of society (i.e. requiring obedience to government edicts or treating citizen equally), which in turn justifies legal rules and systems as a means for enforcing certain constraints that require public (in the sense of a state or government) management and enforcement. Starting with empirical and descriptive assumptions about agents and their environments, constraint can be justified by the inability of any individual to enforce his will on all other agents, by the need for assistance from others to achieve particular outcomes and by the increase and creation of otherwise unobtainable benefits. Rational agents ought to – in the normative sense – voluntarily constrain their actions to bring about a co-operative scheme beneficial to all participants.

This analysis of law uses the methodology of rational choice to validate legal rules. This methodology takes preference-maximisation as the benchmark by which the efficacy of legal rules in promoting preference fulfilment can be judged – these rules are in turn justified because they promote preference fulfilment to a greater extent than any alternative set of rules, lack of rules or different behavioural patterns. The common law criminalisation of certain types of killing makes rational sense because the alternative system of retaliation or duelling are unlikely to as effective due to the higher the detection costs of crimes and criminals on an individual basis. This methodology also explains why the scepticism that moral error theory introduces undermines the validity of moral claims for justifying law. Simply put, it is irrational for agents to constrain themselves in accordance with a set of rules that they consider to be false and which prevent them from maximising their preferences. As morality cannot be shown to be true, agents need only take account of moral claims insofar as

they affect interaction between agents. Rational constraint theory dictates that agents will only accept constraint when this increases their preference fulfilment and when the costs of compliance are not greater than the benefit. Consequently, constraint cannot reflect one moral perspective but must accommodate them all as far as is needed to maintain co-operation.

This philosophical theory of constraint provides a justification for political discourse regarding the distribution of power, control and authority amongst agents while legal rules are justified on the basis of enforcing and maintaining the political co-operative scheme. There are particular concepts used in Hart's and Fuller's legal theories that closely connect with the contractarian tradition – explicitly in Hart's theory and obliquely in Fuller's. The conditions of life for agents and the role of mutual benefit in vindicating constraint as a rational method for co-operation confer *normativity* on legal rules. Agents should follow legal rules because they promote, secure and maintain a co-operative scheme of behavioural norms that allow agents to pursue the realisation of their preferences. However, this also explains why agents should agitate for alternative legal rules when those rules are more efficient and optimal. The following chapters aim to justify and demonstrate the application of this nascent theory of law.

Summary of Articles

A Contractarian Theory of Rational Constraint

Ultimately, this thesis is about justifying behavioural constraints given a numerous, anonymous, heterogeneous population using the theory of rational constraint based on mutual benefit and rational choice. The chapters which constitute this thesis justify and apply this theory of law – chapters one and two explain why we cannot use moral claims and why we should use rational constraint to justify legal rules, while the third chapter explores the implications of this approach for the current legal regulation of reproduction in human rights law and the final two chapters seek to show why moral claims are unable to resolve conflict between agents.

This thesis makes an original and innovative contribution to legal research because it uses a methodology for constructing, justifying and applying legal rules that does not rely on any foundational moral assumptions. Moreover, the individual agent is placed at the core of this theoretical approach as an active part of the

mechanism which produces legal regulation – they are not, as is common in most legal theories, treated as the passive recipients of a legal order. Furthermore, a means of assessing whether a legal rule is legitimate or not is provided – legitimate (i.e. justified) legal rules are those rules which increase preference fulfilment through the imposition of the minimum constraint necessary to achieve that increase. Conversely, legal rules are illegitimate when they impose costs upon agents without increasing preference fulfilment to a greater extent than the loss those constraints impose or when legal rules introduce pure costs without conferring any benefit at all. The criminalisation of homosexuality imposes a pure cost because no preference increase is realised by that constraint and preference satisfaction is reduced to a level below that which is possible in the absence of any law (assuming that that which is not forbidden is permitted).

Thus we have a legal theory based on empirical descriptive assumptions about agents and their environment that justifies constraint because of the benefits that accrue from and are created by co-operation. The law plays a role in ensuring that co-operation *is* the rational choice for agents by attaching penalties to defection from the system but crucially this must be in the interests of agents themselves and is justified by the need to maintain the co-operative scheme. This provides a method for critiquing the historically derived morally infested legal order leading to a more flexible and beneficial legal system. But this is just the beginning of a greater research project that seeks to establish both a political and legal theory of constraint which can mediate between agents with different moral systems and preferences. As human society becomes more complex, more diversified and more disparate conformity to established behavioural constraints deteriorates, conflict increases between distinct segments of society and common ground for agreement diminishes. The social world becomes more fragmented, mutable and unknowable – rationality provides a thread that guides us through a labyrinthine modern society.

Structure of the Thesis

The following chapters provide the structure of this thesis and form the arguments for seeking to develop a non-moral method of analysing and designing legal rules for contemporary populations. The thesis falls in three main parts. Part 1 containing chapters 1-4 explicates the theoretical framework. Part 2 consists of two already published papers, and Part 3 identifies some still outstanding questions, and provides an overall conclusion.

Chapter one begins by arguing that cognitivist objective morality is not truth-apt – that is, either cognitivist morality is false or neither true nor false – because it requires ascribing to the world features that cannot exist independently of those making the observation. Consequently, cognitivist claims can only be asserted and so any agent can refute these claims. This forms the first plank in the overall argument because it motivates us to find a non-moral methodology for regulating interaction between agents. According to the arguments of the first chapter moral claims cannot provide a solution to interaction problems because it will lead to deadlock between proponents of different moralities. Thus we need a method outside of these systems to resolve and regulation inter-agent conduct.

Chapter two takes the next step by proposing a non-moral system that can regulate interaction between self-interested agents by linking the constraints on their conduct to the benefits they receive from others likewise accepting constraint. This system is predicated on the conception of agents as seeking to maximise their preferences and, due to human fragility and vulnerability, this will require co-operation from other agents in nearly all cases. This second chapter therefore analyses the conditions under which an agent should comply with constraint and those circumstances under which an agent may defect from the system by ignoring or violating constraint. In order to reduce the occurrence of defection I propose the inclusion of second-order constraints which can affect different instances of interaction and so increase co-operation. I then use the notion of second-order constraint to explain and justify the existence of legal systems. Together this forms the system I term rational constraint.

Chapter three applies this analysis to human rights law and reproduction arguing that the most rational system will permit the widest range of actions compatible with the maintenance of the system of constraint (in this case human

rights constraint). I conclude that human rights law can permit agents to waive their rights in relation to reproduction in some cases – incidentally simplifying the law – without undermining the system of human rights. This demonstrates the potential for this approach to critique current legal orthodoxy and present alternative legal rules. Chapters four and five seek to demonstrate the problem with using a cognitivist approach to legal regulation that applies to all agents. I challenge the current legal practice regulating access to reproductive technology and its reliance on the assumption that potential persons can be harmed by the actions that create them. By exploring the problems with this approach and the implications of using a person-affecting notion of harm to regulation I aim to show that cognitivist claims cannot be imposed upon rational agents who can reject them (which in turn would reject the regulatory regime premised upon those claims). Thus, we should reject the current legal regime in favour of the rational constraint approach.

In totality, this thesis proceeds by a series of steps. First, I argue that cognitivist moral claims cannot apply to an entire heterogeneous population of agents. Second, we need a non-moral system of rational constraint to determine constraints generally and law in particular. Third, applying this analysis to legal systems and suggesting alternative more rational legal rules. And lastly, that legal constraints that rely on cognitivist claims can be successfully challenged and undermined and thus any agent can – in principle – reject them. All of these steps lead to the proposed new alternative methodology and system of rational constraint which takes the preferences of agents and their conduct as the benchmark for determining the most rational and best laws possible.

Part 1

Chapter 1

Can Morality Justify Population-wide Constraint?

Introduction

Moral systems claim a normative hold over agents and their actions. A normative moral rule is prescriptive – it commands agents to behave in a particular way, it is objective – it is a fixed feature of the environment that agents operate in and it is mind-independent – it requires neither the existence of agents nor their knowledge for validity. Consequently, if such cognitivist realist morality exists this would determine the behavioural constraints that agents should adopt. But this account of morality requires that moral claims are irreducibly normative – moral claims constrain our behaviour because they are *moral claims* – they are the reason for themselves and for our obedience to them. A cognitivist account implies that a method for determining the correct moral system exists which will provide guidance for agents in all circumstances. Moral error theory challenges this conception of moral properties. It holds that ‘ordinary speakers are systematically mistaken about whether there are mind-independent moral facts and properties’¹³⁴ because they ‘assume the existence of properties that cannot exist’.¹³⁵ This anti-realist metaethics precludes cognitivist moral claims, thus we are unable to treat moral claims as truth-apt. The remaining valid constructions of morality cannot provide a single unitary correct set of constraints that all agents can abide by and leaves us with multiple non-truth-apt moral systems. Consequently, we need a means of creating a generally applicable system of normative constraint that is *not* dependent on moral claims.

It is important to note that moral error theory does not imply that moral claims cannot be better or worse – obviously incoherent, contradictory or illogical claims are worse than consistent, coherent and logical claims. Thus moral error theory does not argue against moral reasoning, rather it argues that moral argument is not a fruitful endeavour for resolving conflict between agents with different (but, possibly, equally

¹³⁴ Jonas Olson, *Moral Error Theory* (Oxford University Press 2014) 81.

¹³⁵ Simon Kirchin, *Metaethics* (Palgrave Macmillan 2012) 83.

well reasoned) moral beliefs because moral claims are not truth-apt or are uniformly false. Thus it is not the case that moral claims can be right or wrong rather they can be better or worse, more or less consistent, logical or illogical – consequently, while moral argument might be internally sound it cannot be right, it cannot be a true representation of the world and so it cannot resolve conflict between agents with different moral systems. The fundamental argument of moral error theory is that mind-independent moral claims are metaphysically queer facts completely unlike other provable facts that we have, thus mind-independent moral claims cannot be shown to objectively exist and are false. Consequently, moral claims cannot be truth-apt and so, no matter how well reasoned or argued, cannot conclusively justify population wide constraint because moral claims are, to put it simply, not true for all agents. In a population of homogeneous agents moral systems would be suitable solution to conflict because the entire population would all hold the same set of beliefs but in a heterogeneous population morality fails in its prescriptive function. The argument of metaphysical queerness rules out the kind of cognitivist morality needed to bind all agents because it highlights the mind-dependent mental component of moral claims in constructing moral perspectives. I also look at alternative conceptions of moral claims that can survive the onslaught of moral error theory but these accounts of morality can only survive because they reject the mind-independence crucial to holding all agents to one moral standard. While these non-cognitive accounts can maintain the validity of moral claims it does so only by treating moral claims as relative to the individual holding those views. This chapter demonstrates and concludes that morality – for conceptual reasons internal to moral claims themselves – cannot resolve conflict in heterogeneous populations and so an alternative methodology is needed.

Moral Realism and Moral Anti-realism

An important distinction in metaethics is between moral realism and moral anti-realism, which is linked to the cognitivist and non-cognitivist accounts of morality. Moral realism is the idea that ‘a moral property should inhere *ineliminably* – and thereby, it is safe to assume, *irreducibly* – in the nature of moral experience’.¹³⁶ For

¹³⁶ Iain Brassington, *Truth and Normativity* (Ashgate Publishing Company 2007) 38: Although note that Brassington refers to realism as ‘independentism’, primarily I think, for structural purposes of his book.

example, we come to realise that all acts of torture are morally wrong because these actions have a property of wrongness. Moral realists ‘often embrace two further ideas: cognitivism and descriptivism’.¹³⁷ Cognitivism holds that ‘moral judgements or beliefs are (wholly or primarily) representing states’ of the way the world is – ‘there is some moral stuff and our judgements are attempts to represent it correctly’.¹³⁸ Cognitivists thus hold that moral claims ‘are *truth-apt*, or apt to be assessed in terms of truth and falsity’ because a moral belief is an attempt to accurately represent to ourselves some fact or property about the world.¹³⁹ This leads to descriptivism about moral language – ‘moral language’s (whole or prime) function is to describe stuff in the world’.¹⁴⁰ Do *x* because it is the right thing to do, do not do *y* because that is the wrong thing to do. Most importantly, this means that realism is committed to the claim that ‘an account of the world that hypothesises real moral characteristics is better at explaining the world as it is currently understood than is the best of the other alternatives’.¹⁴¹ If moral claims are true then we can require that agents conform to them because that is the way the world is.

In contrast moral anti-realists (also called moral error theorists) hold that moral properties do not exist, although they are ‘typically, cognitivists and descriptivists’.¹⁴² Anti-realists accept that moral properties should satisfy the conditions of cognitivism and descriptivism because ‘everyday moral judgements are representations of something and typically that ‘something’ is assumed to be moral reality of some sort’.¹⁴³ Cognitivist moral properties should satisfy these requirements because they are statements about how the world is – to claim that a particular act of killing is wrong is to speak of a property that inheres in the action. Moral realists ‘hold that moral judgements are assertions that attribute mind-independent (but non-instantiated) moral properties’ but because these properties are non-instantiated common everyday ‘moral judgements are systematically mistaken’.¹⁴⁴ This systematic error applies to all moral claims that are cognitivist and descriptivist thus there are *no true* moral claims.

¹³⁷ Kirchin (n2) 9.

¹³⁸ Ibid 9.

¹³⁹ Alexander Miller, *Contemporary Metaethics* (Polity Press 2013) 3.

¹⁴⁰ Kirchin (n2) 9.

¹⁴¹ Brassington 32.

¹⁴² Kirchin (n2) 10.

¹⁴³ Ibid 10.

¹⁴⁴ Olson (n1) 4.

Non-cognitivist accounts – which hold that moral claims are reports of mental states, not cognitivist or descriptivist representations of the world – are also anti-realist. Non-cognitivists ‘think that moral judgements express non-cognitive states like emotions or desires’ – for non-cognitivists moral claims are mind-dependent and neither true nor false thus moral claims are not truth-apt.¹⁴⁵ Moral error theory is therefore compatible with – although it does not entail – non-cognitivism. Although non-cognitivists ‘agree with error theorists that there are no moral properties’ they differ from them in holding that moral language is not mistaken.¹⁴⁶ For non-cognitivism moral language *does not* serve to describe the world rather moral language is a means to ‘express our desires and attitudes towards things’ including our moral attitudes.¹⁴⁷ It *can* be true or false whether a particular person has a particular moral attitude or not but the attitudes themselves cannot be true or false.¹⁴⁸ Thus non-cognitivists can hold that cognitivist morality is false without condemning moral claims entirely. Whilst a *pure* moral error theorist would hold that all claims to moral properties are false, non-cognitivism denies the existence of moral *properties* as a feature of the world while holding that we can make moral *claims* about the world because our moral language is an expression of our attitude towards the world. Thus one could be an anti-realist with regard to moral properties proposed by cognitivist theories while holding a non-cognitivist view about moral claims.

With these distinctions in mind I argue in the following sections that realism and cognitivism are, if not certainly false, highly dubious while anti-realist and non-cognitivist accounts permit a number of different conflicting moral claims to exist. If correct, the only valid moral systems would be competing discourses within a population, not a conclusive description of the world that justifies a particular system of constraint. Success in reducing morality and moral systems to one among many discursive frameworks depends on moral error theory to eliminate cognitivist morality. This has significant implications for a system of constraint, specifically moral error theory shows that moral claims are limited in justifying constraint for a population of heterogeneous agents because objective agent-independent prescriptive moral claims are false. We must look for a non-moral source of constraint in a post-moral error theory world.

¹⁴⁵ Miller (n6) 3.

¹⁴⁶ Kirchin (n2) 10.

¹⁴⁷ Ibid 11.

¹⁴⁸ Miller (n6) 3.

Cognitivist Theories of Moral Properties

Within realist cognitivism there are two theoretical branches: the naturalist and the non-naturalist. Naturalists hold that moral judgement is ‘rendered true or false by a natural state of affairs’.¹⁴⁹ Naturalist theories further divide into reductionist and non-reductionist forms: reductionism hold that moral properties are analysable in terms of natural facts, while a non-reductionist holds that moral properties *are* natural facts that ‘are analysable in other [natural] terms’.¹⁵⁰ Non-naturalists, by contrast, hold that moral properties are not themselves natural properties nor are they reducible to natural properties rather they are unanalysable *sui generis* properties.

Both natural and non-natural cognitivists are moral realists, both forms of cognitivism hold that ‘there really are moral facts and moral properties and that the existence of these moral facts ... is constitutively independent of human opinion’.¹⁵¹ It is the belief in the existence of real mind-independent moral properties which makes possible forcible conformity to a moral theory because these moral properties are objectively, intrinsically and independently true for all agents at all times. One difficulty for naturalist theories, of either a reductionist or non-reductionist variety, is explaining how natural facts can give rise to normative claims for ‘[n]atural properties ... do not belong to either [an evaluative or normative] camp’.¹⁵² Of what do cognitivist moral properties consist – are they a ‘ghostly normative substance draping itself over some things but not others?’, are they a ‘nonempirical quality that jumps out from special blends of ordinary empirical attributes?’¹⁵³ Or are they some other kind of property? If naturalist accounts are correct then they will explain and justify the normative content of moral properties.

Non-naturalism

Non-naturalism is the most implausible form of cognitivist morality because it requires the notion of a ‘moral sense’ or some other revelatory means of directly knowing what is moral and what is not. In contrast to their naturalist counterparts, non-naturalists hold that ‘[w]e cannot identify moral properties with other sorts of

¹⁴⁹ Ibid 3.

¹⁵⁰ Kirchin (n2) 50.

¹⁵¹ Miller (n6) 4.

¹⁵² Kirchin (n2) 65.

¹⁵³ Loren Lomasky, *Persons, Rights and the Moral Community* (Oxford University Press 1990) 237.

properties, such as natural properties'¹⁵⁴, moral properties must therefore be '*sui generis*' facts about the world.¹⁵⁵ The property of 'moral goodness [for example] is non-natural, simple, and unanalysable'.¹⁵⁶ The difficulty for non-naturalism is to explain how moral properties can be known if they are non-natural properties because they would not be something that can be known as other properties of the world are known. What 'cognitive faculty allows us to access the fact that justice is good'?¹⁵⁷ Non-naturalism leads to some form of *intuitionist* theory that we simply know – perhaps through gut feeling, moral revelation, conscience or some other sense – of a moral property. This raises the further problem of explaining how *sui generis* moral properties can apply to the wide variety of actions that concern morality. We can criticise non-naturalism as follows:

In admitting that normative ethical concepts are irreducible to empirical concepts, we seem to be leaving the way clear for ... the view that statements of value are not controlled by observation, as ordinary empirical properties are, but only by a mysterious 'intellectual intuition'. A feature of this theory ... is that it makes statements of value unverifiable. For it is notorious that what seems intuitively certain to one person may seem doubtful, or even false, to another.¹⁵⁸

Non-naturalist accounts are implausible because they seem to postulate properties of a kind that must be taken on faith or intuition for there is no other way to 'know' what these properties are or that they exist. For these reasons non-naturalism fails to provide a convincing explanation of moral claims.

Non-reductive Naturalism

Non-reductive naturalism, in contrast, seeks to justify moral properties as independent natural properties of the world – the property of moral wrongness is thus a natural property that attaches to particular actions *independently of any other natural properties*. The Cornell Realists exemplify this view. They argue that moral

¹⁵⁴ Kirchin (n2) 7.

¹⁵⁵ Ibid 40.

¹⁵⁶ Miller (n6) 4.

¹⁵⁷ Ibid 32.

¹⁵⁸ Alfred Ayer, '*Language, Truth and Logic*' (2nd edn, Gollancz 1946) 130.

properties exist ‘as part of the natural fabric of the world’ for the same reason that the properties of the physical sciences are real ‘namely, *that they pull their weight in explanatory theories*’.¹⁵⁹ The argument goes that ‘[t]here are various moral phenomena that we notice around us all the time’ – the ‘supposition is that the best explanation of why such things exist is that there exist moral properties that cause them’.¹⁶⁰ Cornell Realism thus holds the following:

(1) P is a real property if and only if P figures ineliminably in the best explanation of experience.

(2) Moral properties figure ineliminably in the best explanation of experience.

Therefore:

(3) Moral properties are real properties.¹⁶¹

This view has been challenged by Gilbert Harman’s best explanation theory which argues that moral properties are not real natural properties because ‘moral “properties” do not figure ineliminably in the best explanation of experience’.¹⁶² This is because ‘[p]ositing value as an objective feature of the warp and woof of the world seems metaphysically extravagant’¹⁶³ and ‘the ascription of moral experience to moral “reality” contributes *nothing at all* to the explanation of that experience’.¹⁶⁴ In other words, we can explain all moral phenomena in terms of agents themselves.

Any description of an action that we find morally unacceptable can eliminate all references to moral properties without suffering any ‘explanatory loss’.¹⁶⁵ Consider the case of someone setting a cat on fire for the pleasure of watching it suffer – we can describe everything that happen in natural terms without reference to moral properties thus moral properties cannot feature *ineliminably* in our best explanation of experience. It might be that moral properties play an explanatory role ‘via some version of reductionism’ because moral properties would then be reducible to natural facts which *do feature ineliminably* in explanations of human experience

¹⁵⁹ Miller (n6) 145.

¹⁶⁰ Kirchin (n2) 51.

¹⁶¹ Miller (n6) 145.

¹⁶² Ibid 145.

¹⁶³ Lomasky (n20) 237.

¹⁶⁴ Brassington (n3) 34.

¹⁶⁵ Miller (n6) 145.

but non-contingent non-reductionist moral properties are eliminable.¹⁶⁶ However, it seems that a better explanation of moral claims comes from the ‘psychology or moral sensibility of the person making the moral observation’.¹⁶⁷ Setting fire to a cat may be wrong because we empathise with sentient beings, because we recognise pain as unpleasant or because we particularly like cats – if moral properties were reducible to these natural psychological facts they would be performing an explanatory function (albeit not an essential one as it is psychology which is actually explaining things).

Non-reductionist naturalism attempts to challenge this claim by providing an account of moral properties in natural terms without reducing moral properties to natural properties. Returning to the example of someone setting a cat on fire, we might argue that ‘an act which possesses the non-moral property of *pointless, deliberate cruelty*, ... or some other natural properties upon which moral properties supervene, also possesses the moral property of *wrongness*’.¹⁶⁸ The argument goes: ‘*it is simply false to say that if the act had lacked the property of pointless, deliberate cruelty, we would still have believed the action to be wrong*’.¹⁶⁹ If “we find it plausible to attribute causal efficacy and explanatory relevance to moral facts, [then] why should we not conclude’¹⁷⁰ that moral properties exist?

Yet even if we accept that wrongness supervenes in this manner, ‘a moral epiphenomenalist can accept the relevant counterfactual judgement without having to suppose that moral features of actions ever explain any nonmoral facts’.¹⁷¹ In the cat example, a moral epiphenomenalist could accept the role the natural fact of pointless deliberate cruelty plays in explaining the wrongness of the act without needing to postulate an *additional* moral feature of wrongness. In other words, the feature of pointless cruelty is part of the explanation of what has happened but there is no need to take these features as causing a further property of moral wrongness. On the non-reductionist account, moral wrong would simply be an additional property that is present but not one that is casually linked to any others. Similarly, in an explanation of Adolf Hitler’s behaviour we need to ask ‘whether our claim that Hitler was

¹⁶⁶ Ibid 147.

¹⁶⁷ Gilbert Harman, *The Nature of Morality* (Oxford University Press 1977) 6.

¹⁶⁸ Miller (n6) 150.

¹⁶⁹ Ibid 150.

¹⁷⁰ Nicholas Sturgeon, ‘Harman on moral explanations of natural facts’ (1986) 24 *Southern Journal of Philosophy* Supplement 75.

¹⁷¹ Gilbert Harman, ‘Moral explanations of natural facts – can moral claims be tested against reality?’ (1986) 24 *Southern Journal of Philosophy* Supplement 63.

depraved is most adequately described by an appeal to moral reality, or by an appeal to psycho-social facts'.¹⁷² Arguably the latter is a better explanation because 'psycho-social facts are independently testable in a way that, contrary to the implicit demands of the RI¹⁷³ [realist] theorist, moral facts are not'.¹⁷⁴

Non-reductionist naturalism fails to explain why moral properties (as separate distinct natural properties) are needed to explain the wrongness or rightness of an action – 'positing a real moral characteristic adds nothing to the explanatory story'.¹⁷⁵ Moral properties cannot be treated as facts in the same way as facts from natural science. While natural science is the best explanation of observable phenomena, 'real moral characteristics do not feature in the best explanation of observed phenomena because they do not really feature in *any* explanation of observed phenomena'.¹⁷⁶ Given this, one might be tempted to say that pointless cruelty *itself* is wrong, that the feature of pointless cruelty or the harm caused comprises wrongness rather than introducing a new separate moral property. This alternative approach is suggestive of reductive naturalism because it treats moral facts as reducible to natural facts.

Reductive Naturalism

Reductive naturalism tries to account for the 'one-to-one and one-to-many mappings' between moral and natural properties by treating moral properties as reducible to a natural property or properties.¹⁷⁷ One proposal, from Peter Railton, suggests 'that moral properties are reducible to complex Natural properties, via a synthetic property identity delivered courtesy of an empirically justifiable reforming definition'.¹⁷⁸ By using this programme moral properties can 'be interpreted as standing for some naturalistic property N, and [we] then investigate whether N contributes to the *a posteriori* explanation of features of our experience'.¹⁷⁹ This is done by combining a *full information analysis*¹⁸⁰ with a social point of view, which is the key feature of moral properties. The full information analysis works as follows:

¹⁷² Brassington (n3) 39.

¹⁷³ RI refers to the 'realist independentist', who holds that moral facts "just *are* right or wrong, good or bad by virtue of some characteristics or property that they possess in and of themselves": Ibid 28.

¹⁷⁴ Ibid 39.

¹⁷⁵ Ibid 44.

¹⁷⁶ Ibid 38.

¹⁷⁷ Kirchin (n2) 69.

¹⁷⁸ Miller (n2) 184.

¹⁷⁹ Ibid 182.

¹⁸⁰ The term 'full information analysis' is taken from Miller's *Contemporary Metaethics*.

[Consider] A+, who has complete and vivid knowledge of himself and his environment, and whose instrumental rationality is in no way defective. We now ask A+ to tell not what *he* currently wants, but what he would want his non-idealised self A to want ... were he to find himself in the actual condition and circumstances of A.¹⁸¹

Full information analysis thus takes the facts of an individual's situation as determining their good – although at this point 'good' is characterised in non-moral terms. Consider Lonnie (an example from Railton), he is dehydrated and has a desire for a glass of milk which will make him feel worse although he is unaware of this fact. Under the full information analysis programme Lonnie+, who is fully informed and instrumentally rational, would want the non-ideal Lonnie to form a desire for a glass of water and lose his desire for a glass of milk. Consequently, the circumstances and constitution of Lonnie determine what his good is and these are natural facts of the situation in which he finds himself.

The next step is to extend the use the full information analysis to provide an account of moral properties. Railton treats morality as 'concerned most centrally with the assessment of conduct or character where the interests of more than one individual are at stake'.¹⁸² In order to successfully consider the interests of multiple individuals 'philosophers seek to capture the special character of moral evaluation by identifying a *moral point of view* that is impartial, but equally concerned with all those potentially affected'.¹⁸³ This moral 'social point of view' consists of 'what would be rationally approved of were the interests of all potentially affected individuals counted equally under circumstances of full and vivid information'.¹⁸⁴ Therefore, '*x is morally right if and only if x would be approved by an ideally instrumentally rational and fully informed agent*' who equitably represents all agents.¹⁸⁵ A good example of this might be the health benefits of vaccination – if all agents would ideally want to be vaccinated then this suggests that vaccinations are morally good for all agents. Full information analysis defines moral properties in terms of ideally rational decisions

¹⁸¹ Peter Railton 'Moral Realism' (1986) 95 *Philosophical Review* 163, 173-4.

¹⁸² Ibid 189.

¹⁸³ Ibid 189.

¹⁸⁴ Ibid 190.

¹⁸⁵ Miller (n6) 196.

about non-moral goodness thus moral properties are real explanatory facts reducible – through a revisionist *a posteriori* analysis – to natural properties.

The full information analysis rests upon the hypothetical comprehension of this ideally rational perspective. But this perspective can be challenged on the basis that *we* cannot – those actual people making the decision – cannot ever have enough information to even begin to approximate the necessary level of information for the analysis. Unless we can comprehend, to some degree at least, the ideal position the full information analysis will not provide any guidance for agents. This challenge focuses attention on the heart of full information analysis – that it should be practically useful in identifying moral properties. But the ideal self would have to ‘be the same agent who experiences first-hand each type of life ...[we] might live’ otherwise we would be unable to determine how and what kind of life we should lead.¹⁸⁶ They would have to be everyone who we could possibly be (assuming this is a coherent notion) in order to be able to make a fully informed judgement about the moral properties of each life. In other words, we would need to have full information not just about our actual life but about all our possible lives in order to be able to judge what is best for us. Even if we restricted our consideration to potential lives we might lead from now, from this moment onwards, the information costs would still be insurmountable. This is shown by trying to use two hypothetical methods that this experiential A+ could take: they could consider each life serially or they could live each life and then forget it until they had lived all lives and then remember them all. These two methods are the serial method and the amnesia method respectively.

We must, however, reject both of these two methods as providing a solution to the full information account. In the first case, any account of lives lived serially would distort the perception of subsequent lives. Consider belief or not in a god or gods: if ‘it is a fact that God does not exist, and we are given this knowledge’ as idealised agents, then we would be ‘unable to experience adequately what it would be like for us to be born-again Christians’.¹⁸⁷ The alternative amnesia model, whereby our fully informed self could live each life without the distorting effect of mutually exclusive and incompatible experiences by suddenly recalling all lives simultaneously, fails to resolve this problem. The amnesia model would be unable to determine which lives were to be preferred because each life will be judged from an internal viewpoint thus

¹⁸⁶ David Sobel, ‘Full Information Accounts of Well-being’ (1994) 104 *Ethics* 784, 801.

¹⁸⁷ *Ibid* 802.

raising the problem of interpersonal comparison (interpersonal because each life is considered separate from the others). The fundamental basis of the objection to the full information analysis is that it can only work successfully if A+ exists as a single, unifying perspective across all lives and can commensurably order the multiplicity of life options. But ‘we do not have a single informed perspective to deal with, but several and each will offer conflicting assessments of where the agent's well-being lies’.¹⁸⁸ Even if we could actually imagine living different lives this would not provide us with access to a moral point of view that would be able to order all the options open to us.

Alexander Miller suggests an alternative solution.¹⁸⁹ He proposes that a distinction be made between cases where the idealised self suggests that pursuit of Q at time t and the pursuit of R at time t^* where Q and R are compatible and incompatible. That is either one could pursue Q for a time and then pursue R without either undermining the pursuit of the other or the pursuit of Q makes the pursuit of R impossible. Distinguishing these two cases allows us to argue that compatible cases are unproblematic while incompatible cases are impossible. As idealised selves ‘are both fully informed and ideal in point of instrumental rationality ... [the temporally prior idealised self] can be assumed to know’ what will be wanted at both time t and time t^* and thus cannot (due to instrumental rationality without defect) choose a course of action that would frustrate future action.¹⁹⁰ For example, presumably the idealised self would, at no point, want the non-idealised self to believe in a god if the non-existence of gods is a fact.

The full information analysis attempt to explain the mapping relations between moral and natural properties is unsuccessful because it cannot provide us with the necessary experience or knowledge for explaining moral properties. If, however, such a programme were successful it would provide us with the normative force necessary for following moral prescriptions because the recognition of natural facts would *directly* give rise to moral courses of action. The difficulty for all naturalist cognitivist realists is in explaining how moral properties can exist as mind-independent real properties of the world and the process through which we can correctly know that these moral claims exist and what they prescribe given that such properties seem to be

¹⁸⁸ Ibid 805.

¹⁸⁹ Miller (n6) 212.

¹⁹⁰ Ibid 212.

incorporeal unobservable phenomena. As we have seen moral properties can be eliminated from explanations of observable phenomena while appeals to idealised states of knowledge present an insurmountable obstacle to making practical decision under the condition of limited information available to us. In short, the fundamental problem for cognitivist theories is the metaphysical queerness of their hypothesised moral properties.

Moral Error Theory and the Arguments from Queerness

John Mackie's anti-realist moral error theory argues that if 'moral judgements are apt to be true or false ... moral judgements are in fact always *false*'.¹⁹¹ Mackie argues that 'if there were objective values [real moral properties], then they would be entities or qualities or relations of a very strange sort utterly different from anything else in the universe' – they are features of the world inexplicable using the same method by which we acknowledge physical reality.¹⁹² Jonas Olson refers to this as *standard error theory* in which moral claims either 'are uniformly false or ... they are uniformly neither true nor false'.¹⁹³ *Moderate error theory*, in contrast, holds that '[w]e might be systematically mistaken about the nature of moral properties and facts, but there might nevertheless be moral properties and facts'.¹⁹⁴ To access these facts would require 'some special faculty of moral perception or intuition, utterly different from our ordinary ways of knowing everything else'.¹⁹⁵ The difficulties faced by cognitivist accounts are put front and centre in moral error theory – rather than attempt to circumvent this problem, moral error theory uses it as the basis for the argument against the entire cognitivist programme. Yet it is the following further characteristic that cognitivist morality requires which confers great force to moral error theory. Cognitivist 'objective good [a moral claim] would have to be sought by anyone who was acquainted with it, not because of any contingent fact [about this agent or every agent] ... but because the end has to-be-pursuedness somehow built into it'.¹⁹⁶ Taken together with the problems cognitivists face, this forms the backbone of argument from queerness – moral properties are *metaphysically queer* because 'our concept of a moral fact is a concept of an *objectively prescriptive* fact' but 'there are

¹⁹¹ Miller (n6) 5.

¹⁹² John Mackie, *Ethics: Inventing Right and Wrong* (Penguin Books 1990) 38.

¹⁹³ Olson (n1) 9.

¹⁹⁴ *Ibid* 9.

¹⁹⁵ Mackie (n59) 38.

¹⁹⁶ *Ibid* 40.

no objectively prescriptive facts or properties'.¹⁹⁷ This is because we impose upon the world values, properties and facts – we might rank different metals according to our beliefs and values but gold has no more objective value than iron pyrite. Moral error theory argues that moral claims are no different to the value assigned to gold over iron pyrite. We should thus conclude that the 'assertion that there are objective values or intrinsically prescriptive entities ... [is] not meaningless but false'¹⁹⁸. Thus '(positive, atomic) moral judgements are uniformly false: our moral thinking involves us in a radical *error*'.¹⁹⁹ Consequently, requiring agents to conform to moral claims loses a great deal of force because agents must play some originating role in the only remaining kind of morality – anti-realist morality.

But what exactly does it mean to say that our conception of a moral fact or property is *objectively prescriptive*? What does *to-be-pursuedness* mean? How do these phrases relate to moral statements? Olson usefully parses the queerness argument into four varieties, which differs from Mackie who identified only two components to the argument from queerness – 'one metaphysical, the other epistemological'.²⁰⁰ The most important refinement of the original metaphysical claim is that morality 'entail[s] irreducibly normative reasons'.²⁰¹ The argument from queerness based upon the claim that the irreducibly normative relation between moral claims and agents is metaphysically queer (that is, non-real) is the most successful form of the argument from queerness. There are, however, a number of arguments from queerness – Mackie's metaphysical and epistemological arguments from queerness and Olson's parsing of Mackie's arguments into 'four distinct queerness arguments, focusing on supervenience, knowledge, motivation and irreducible normativity'.²⁰² These four arguments focus on different elements of moral realism. The first, concerns claims 'that the moral supervenes on the natural'.²⁰³ The second looks at the epistemological problem 'that we do not (and perhaps cannot) gain knowledge about moral facts and properties'.²⁰⁴ The third motivational argument

¹⁹⁷ Ibid 105.

¹⁹⁸ Ibid 40.

¹⁹⁹ Miller (n6) 5.

²⁰⁰ Mackie (n59) 38.

²⁰¹ Olson (n1) 117.

²⁰² Ibid 79.

²⁰³ Ibid 88.

²⁰⁴ Ibid 100.

relates to the purported motivation-inducing effects of moral claims and, finally, the fourth argument challenges the irreducible normative function of moral claims.

Unfortunately, the first three arguments all fail and only the fourth argument from normative queerness is successful. The supervenience argument fails because it misfires by targeting non-naturalist but not naturalist cognitivism. By claiming that supervenience is metaphysically queer this argument ‘is targeting the non-naturalist view that every instance of a moral property is an instance of a non-natural property that obtains in virtue of a distinct natural property’ because it holds that moral properties cannot appear out of nowhere and supervene upon natural properties.²⁰⁵ However, the argument from supervenience does not refute the claim that moral properties *are* natural properties – that the property of wrongness is a natural feature of certain actions. The epistemological argument from knowledge fails because ‘moral knowledge is or requires synthetic *a priori* knowledge and if there are other examples of synthetic *a priori* knowledge, it is not clear that the moral non-naturalists’ account of moral epistemology is *epistemologically* less simple or less comprehensible than alternative accounts’.²⁰⁶ If we can have synthetic *a priori* knowledge that ‘[n]othing can be red and green all over at the same time’ or ‘the laws and inference rules of logic, such as the law of non-contradiction and *modus ponens*’, it is at least possible that we could have such knowledge about moral properties.²⁰⁷ The argument from epistemology targets, therefore, the possibility of synthetic *a priori* knowledge generally which would leave us without a functioning conceptual framework at all rather than removing a false conception of moral claims. Thus this argument ‘fail[s] to isolate uniquely queer features of moral and normative properties and facts’ resulting in a scatter shot attack on epistemology generally.²⁰⁸

Finally, the argument from motivation fails because it ‘presupposes a necessary connection between first-hand knowledge of moral facts and motivation to act accordingly’ – that is the argument from motivation presupposes an internalist account of moral facts.²⁰⁹ However, as the internalist/externalist debate in philosophy shows this is by no means necessary for moral facts to be true, thus an externalist account can neutralise the argument from motivation. Nevertheless, the motivational

²⁰⁵ Ibid 99-100.

²⁰⁶ Ibid 103.

²⁰⁷ Ibid 102.

²⁰⁸ Ibid 103.

²⁰⁹ Ibid 111-2.

argument will be discussed here in more detail (along with the successful argument from irreducible normativity) because the discussion of motivation will inform our notion of normativity making the content of the queerness argument from irreducible normativity clearer.

The Argument from Motivation

For moral properties to be ‘intrinsically prescriptive entities’ morality must have ‘to-be-pursuedness somehow built into it. Similarly, if there were objective principles of right and wrong, any wrong (possible) course of action would have not-be-doneness somehow built into it’.²¹⁰ If, according to this argument, cognitivist morality is right then moral claims would have to possess certain inherent features that move an agent to action from *mere comprehension* of the moral claim. Olson suggests that there are two possible arguments based upon this statement – but ‘[u]nfortunately Mackie does not clearly distinguish them’.²¹¹

On the one hand, moral facts are or entail demands that agents act in certain ways. This is the *normative* relation. On the other hand, moral facts motivate anyone who knows about them, or is acquainted with them, to act in accordance with them. This is the *psychological* relation.²¹²

The notion of *to-be-pursuedness* indicates that moral claims should possess both the normative and psychological relation. As a consequence of the internalist and externalist debate in metaethical theory ‘most philosophers nowadays distinguish sharply between *normative reasons* and *motivating reasons*’.²¹³ Mackie’s analysis entails a strong version of motivational internalism for it implies that mere knowledge of moral properties would *necessarily* result in an agent being motivated to act in accordance with that property.

But the claim that prescriptive mind-independent moral claims are representational cognitive beliefs about the world means that, as representations of the world, they cannot have this motivational force. This suggests that ‘what he [Mackie] finds queer is rather that moral facts would be such as to *guarantee* motivation in

²¹⁰ Mackie (n59) 40.

²¹¹ Olson (n1) 104.

²¹² Ibid 104.

²¹³ Ibid 105.

anyone acquainted with them'.²¹⁴ This strong interpretation of the argument from motivation is unwarranted because moral claims could be representational beliefs that provide weak motivation to act in a certain way *without* entailing that any and all agents would be compelled to act in that way – although moral claims would lose a great deal of their power to induce certain behaviour on this weaker interpretation. The motivational argument can be challenged by the claim that 'such [moral] facts exert motivational pull on anyone acquainted with them' without being sufficient to necessarily entail that an agent acts in accordance with its prescriptions.²¹⁵

The argument from motivation, therefore, fails because it 'presupposes a necessary connection between first-hand knowledge of moral facts and motivation to act accordingly'.²¹⁶ This fails to explain 'amoralism and weakness of will' both of which 'allow for the possibility that people sometimes judge, and even know, first-hand that some action is wrong without being at all motivated not to do it'.²¹⁷ These phenomena lead some to conclude that internalism is false and that, in addition to a belief about moral claims, a desire to be or act morally is also necessary for someone to conform to moral claims. This externalist view, in which motivation is not a conceptual necessity of moral properties,²¹⁸ counters the argument from motivation if it is premised on an internalist understanding of moral properties rendering the motivational queerness argument ineffective. The move to externalism is possible because a cognitivist can claim that 'we can fail to be motivated by true or false judgements about objective and non-conventional, or irreducibly normative, reasons and requirements'.²¹⁹ Thus the motivation argument fails to explain why morality is metaphysically queer because motivation can be pulled apart from the normative demand of moral claims.

Motivational queerness disappears once these two concepts are separated but it does help highlight that the idea of objectivity implies that moral claims 'are

²¹⁴ Ibid 108-9.

²¹⁵ Ibid 109.

²¹⁶ Ibid 111-2.

²¹⁷ Ibid 113.

²¹⁸ Externalism is also referred to as the Humean theory of motivation. We are not concerned with the substantive internalist/externalist debate, only with how these two conceptions of motivation are relevant to the motivational queerness argument. For discussions of the substantive issues see Simon Kirchin, *Metaethics* (Palgrave Macmillan 2012) Chapter 7, Michael Smith, *The Moral Problem* (Blackwell Publishing 1994) Chapters 4 and 5, Alexander Miller, *Contemporary Metaethics* (Polity Press 2013) s.10.4 pp.265-73 and Andrew Fisher and S. Kirchin (eds), *Arguing About Metaethics* (Routledge 2006) Parts 10 and 11.

²¹⁹ Ibid 114.

normative in the sense that they entail that there are facts that favour certain courses of behaviour, where the favouring relation is irreducibly normative'.²²⁰ Even if motivation is separable from moral claims, normativity is not because morality is supposed to provide reasons for behaving in certain ways that are independent of the desires of agents themselves. The cognitivist moral prescription against torture, for example, is not based on whether an agent dislikes torture, rather the claim is that no agents should torture anyone *even when they enjoy doing so*. Thus it seems that motivational pull should come from the normative favouring relation rather than from some psychological desire for normativity is the core of morality's claim to command, proscribe and prescribe certain behaviours. If the motivational queerness argument boils down to the claim that normativity provides motivation then it is the notion of normativity in use which must metaphysically queer because it is this notion that requires agents conform to moral prescriptions. For this reason it is the argument of irreducible normativity – not the argument from motivation – that successfully challenges our cognitivist conception of moral claims as providing objective mind-independent behavioural prescriptions for all agents.

The Argument from Irreducibly Normativity

The best understanding of cognitivist moral claims is that they 'entail irreducibly normative reasons' because this captures the sense that moral claims are reasons in and of themselves which apply to and bind all agents requiring them to act in certain ways.²²¹ Strictly speaking, 'it is the irreducibly normative favouring *relation* [between a reason and the behaviour it prescribes] ... that is queer' not the reason itself.²²² Consider Olson's example: 'the fact that one ought morally to donate 10% of one's income to Oxfam entails that there is a reason to do so. The reason might be the fact that donating to Oxfam promotes human well-being'.²²³ However, the favouring relation between the action 'donate to Oxfam' and the promotion of human well-being is irreducibly normative because the reason for donation is 'not contingent on the agents' desires'.²²⁴ 'Whether or not agents desire to promote human ... well-being ... they have moral reasons ... to donate 10% of their income to Oxfam' because

²²⁰ Ibid 114.

²²¹ Ibid 117.

²²² Ibid 118.

²²³ Ibid 118-9.

²²⁴ Ibid 119.

irreducible normativity entails that '[o]ne cannot escape moral reasons by adverting to one's desires'.²²⁵ We can see that moral claims do not require that agents want to be moral, they simply require that agents comply with them because moral claims are normative reasons for action – it is this relation that moral error (in its more refined state) argues is false because such a relation between reason and agents is metaphysically queer.

Irreducible normativity is further distinguishable from the use of normative as meaning 'to do with norms, rules, or correctness'.²²⁶ Thus '[t]o say that some behaviour is correct or incorrect according to some norm, N, is not to say anything normative'.²²⁷ Normativity, as we are concerned with it, would be the irreducible favouring relation to conform to N, not compliance with the rules that N produces. 'There is nothing metaphysically queer about conventional norms, rules, or standards of correctness that require or recommend various courses of behaviour' because they are tied to psycho-social factors about agents.²²⁸ Thus, 'error theorists can recognise reasons that reduce to facts about agents' *roles* and rule-governed *activities*'.²²⁹ These kinds of reasons are '*reducible* reasons' – they are 'reducible to facts about what promotes desire satisfaction, or correctness norms that may or may not be conventional'.²³⁰ Consequently, moral error does not deny that there may be reason to act according to moral claims but these reasons are contingent, context-dependent, admixed and agent-relative. In contrast, irreducibly normative reasons are 'not reducible ... to facts about what would promote satisfaction of A's desires, or to facts about A's roles or engagement in rule-governed activities'.²³¹ Cognitivist moral claims rely on irreducible normativity to confer force upon them in affecting the behaviour of agents whose reasons, desires and goals may contravene moral codes. Without irreducible normativity moral systems are contingent conceptual frameworks that cannot conclusively determine an agent's behaviour. Moral claims should be obeyed because they are reasons for certain behaviours that are irreducible to any other agent-relative factors.

²²⁵ Ibid 119.

²²⁶ John Broome, 'Is Rationality Normative?' (2007) 2 *Disputatio* 161, 162.

²²⁷ Olson (n1) 120.

²²⁸ Ibid 121.

²²⁹ Ibid 121.

²³⁰ Ibid 121.

²³¹ Ibid 122.

Unfortunately, this is ‘admittedly not very illuminating’ but it is ‘very doubtful that we can get much further in attempting to throw light on the notion of normativity’.²³² Indeed, it seems that irreducibly normativity is defined solely in terms of its opposition to reducible normativity – consequently, Olson takes normativity ‘to be a primitive notion’²³³. What we are trying to grasp is the idea that a moral claim can prescribe or proscribe certain behaviours simply because it is *the moral thing to do* and not due to some other contingent agent- relative factor such as their attitudes or desires. We might best represent the irreducible normativity of cognitivist morality as follows: *you should bring about x*. No reason need be given because moral claims are normative in and of themselves, no further illumination of the concept can be achieved and no argument can be given for the existence of irreducible normativity – one either accepts it or rejects it. The distinction between irreducible and reducible normativity is important because reducible normativity ties into the preferences or desires of agents *giving them* a reason to comply with it. Nevertheless, with this – somewhat sketchy – distinction in mind the queerness argument is now:

Irreducibly normative favouring relations are queer. Hence, moral facts entail queer relations. If moral facts entail queer relations, moral facts are queer. Hence, moral facts are queer.²³⁴

This argument accepts the ‘*conceptual claim*’ that ‘moral facts are or entail irreducibly normative reasons’ but denies the ‘*ontological claim* that there are irreducibly normative reasons in reality’ for they are metaphysically queer entities which are not needed for explaining the world.²³⁵ But because *moral* error theorists reject irreducible normativity they must be error theorists ‘not only about morality but about normativity more generally’ – with one qualification, error theorists need not hold that *reducible* normativity is in error because this is explainable in agent-relative terms.²³⁶ For reducible normativity can operate as follows: if you want *x* then you should do *y* because this brings about the state of the world you are trying to achieve.

²³² Ibid 122.

²³³ Ibid 122 (footnote omitted).

²³⁴ Ibid 124.

²³⁵ Ibid 124.

²³⁶ Ibid 123.

Thus the irreducible normativity of moral claims is the fulcrum upon which moral error theory turns but this also provides the basis for an argument against it.

One counter argument claims that the conceptual claim error theory ‘attributes to ordinary moral discourse an error that simply is not there; ordinary moral claims are not and do not entail claims about irreducible normative reasons’.²³⁷ Moreover, error theory relies on the claim that moral claims have a core, non-negotiable belief in irreducible normativity. This can be challenged on the basis that ‘[b]eliefs important for morality – such as a belief in the existence of moral demands – may have a number of different interpretations and readings’.²³⁸ On this view ‘morality does not seem to have a distinctive belief or small set of beliefs that are non-negotiable’.²³⁹ Thus moral error theory is not wrong but it targets a kind of morality that does not exist – namely, an authoritative morality of objective factual irreducibly normative demands when in fact morality consists of different contingent interpretations. The question is: does moral error theory target a form of morality present within our conceptual framework or does it not? Or, in other words, do people conceive of morality as cognitivist or non-cognitivist?

When we interrogate the reasons why people think certain actions are (im)moral we might conclude that they do not hold moral claims to be absolute facts – but this seems unlikely given the way people speak and the extreme acts that moral systems can induce. It seems clear that when someone states *you should not do that, it is wrong* or *you should do that, it is right* they are using moral claims to induce certain behaviour *regardless* of agent-relative factors. This would only make sense if moral claims were irreducibly normative because even if we say *you should donate to charity, it is a good to help others* the good here is a moral claim. Otherwise we would say you donate to charity because it is beneficial to you, because you have stated an aim of helping others or because you want to improve the lives of others – all of which would give the specific agent being told a reason to act. Moral claims on the other hand are aimed at all agents not just those who would benefit from or have a particular reason to act morally. If moral claims *are* to apply to *all* agents then, contrary to the argument of non-negotiable belief, those claims must be using irreducible normativity because only this removes the need to provide specific reasons

²³⁷ Ibid 126.

²³⁸ Kirchin (n2) 90.

²³⁹ Ibid 91.

for specific agents to act in a certain way. Thus, the ‘most straightforward explanation of why moral claims have the kind of force’ necessary for demanding conformity is ‘*because they are or entail claims about irreducibly normative reasons*’.²⁴⁰

But this raises the further question of why anti-realism is not the dominant notion of morality. I will only offer a speculative answer to this question but one that I think is plausible. In an environment of constant change and uncertainty humans impose rules, guidelines and structure through their mental mapping of the world – we frequently treat things as unchanging despite the vast empirical evidence to the contrary. For example, murder has long been a crime but what constitutes murder has changed significantly over the centuries – if I shot someone in the course of duel in the 18th and 19th centuries I would not be considered a murderer. Despite this, the need to function in everyday circumstances leads to an interpretation of reality that includes fixed rules of behaviour – this, I suggest, leads the dominance of realist tendencies in moral discourse and attitudes. Thus humans are psychologically predisposed to treat their environment as fixed including the rules by which we live and this accounts for the seeming ineffectualness of anti-realist reasoning. Additionally, as we shortly see, there seems to be a general reluctance to countenance the possibility that morality is contingent as this would seem to undermine universal rights and entitlements. For these reasons irreducible normativity plays a crucial role in the commanding nature of moral claims despite the more compelling case of anti-realism.

It is therefore arguable that moral systems require at least one moral claim of irreducible normativity to ground the rest of the system’s claims on the behaviour of all agents. Consider the claim that ‘[t]o have moral standing is to be the sort of being whose interests must be considered from the moral point of view’²⁴¹. Or the claim that ‘we must go beyond a personal or sectional point of view and take into account the interests of all those affected’.²⁴² These notions of the moral point of view seem to be claims to irreducible normativity as they seem to be unattached to any end that agents may have which allows them to appear as fixed immutable rules. Whether or not these moral points of view fulfil the interests, desire, preferences or goals of agents does not

²⁴⁰ Olson (n1) 129.

²⁴¹ Bonnie Steinbock, *Life Before Birth* (Oxford University Press 2011) 1.

²⁴² Peter Singer, *Practical Ethics* (Cambridge University Press 2011) 10.

seem to matter – the reason for adopting the moral point of view seems to be the irreducibly normative reason that it *is* the moral point of view.

Even if moral error theory targets a concept that is absent in everyday morality, this does not undermine the claim that a system of normative ethics must proceed upon the basis that irreducibly normative objective moral claims are impossible. The counter arguments just mentioned do not disagree with the conclusion of moral error theory, at most – if they are right that moral systems do not rely on irreducible normativity – they are claiming that the kind of morality moral error theory argues against has never existed *but this is also the conclusion of moral error theory*. Essentially, the argument would at best show that moral error theory is redundant because morality does not claim to be irreducibly normative – they do not provide an argument for concluding that moral error theory is wrong. However, as stated earlier, it seems that some people do treat moral claims as cognitivist objective facts and that many certainly do so implicitly when condemning certain behaviours.

One final fallback position is available to opponents of moral error theory, namely, they can ‘maintain that it is a fundamental fact about reality that there are irreducibly normative reason relations’ because the question of irreducible normativity ‘is at a bedrock metaphysical level’.²⁴³ In other words, because the irreducible normative relation is a primitive unanalysable assumption, it can be asserted but not demonstrated and so one can simply assert that moral claims are irreducibly normative. But this means that opponents of irreducible normativity can equally assert that there is not such thing – either way conflict is not resolved by recourse to moral analysis. If this is right then Simon Kirchin is perhaps the most honest theorist when he says ‘I find it hard to believe that there are evaluative properties, or prescriptions, or whatever, that are created and continue to exist mind-independently’.²⁴⁴ This then may be the best objection there is to claims to irreducible normativity – irreducible normativity is just, *literally*, unbelievable. Anti-realists (moral error and non-cognitivist theorists) find this relation queer whilst the moral realist seems not to – and if the irreducibly normative favouring relation is an unanalysable metaphysical assumption that can only be asserted – then proponents and opponents of moral error theory are left ‘staring incredulously at each other’²⁴⁵.

²⁴³ Olson (n1) 136.

²⁴⁴ Kirchin (n2) 138.

²⁴⁵ Olson (n1) 136.

Moral error theory forces those who claim that morality is intrinsic, objective, prescriptive, mind-independent and binding on all agents to rely on a metaphysically primitive conception of irreducible normative favouring relations. If this is the case, we can reject moral claims simply by rejecting the asserted primitive notion of irreducible normativity. However, this does not necessarily commit us to either rejecting the possibility of morality entirely, nor to holding that moral claims are entirely subjective or relative. The ‘denial that there are objective values does not commit one to any particular view about what moral statements mean, and certainly not to the view that they are equivalent to subjective reports’.²⁴⁶ This is why we can still have better or worse moral reasoning and argumentation without moral claims ever being truth-apt. Thus while post-moral error theory morality is broadly subjective it does not necessarily entail that moral claims are formed by every individual’s whim and fancy. Moral error theory *does* support the view that moral claims are ‘not to be discovered but to be made’, thus it identifies what kind of moral claims *we cannot have* but does not state what kind of moral claims *we can have*.²⁴⁷ There are a number of accounts of moral claims that can result from and are compatible with moral error theory – the three most prominent theories are abolitionism, fictionalism and non-cognitivism.

Abolitionism and Fictionalism

Abolitionism responds to moral error theory by excluding moral properties that are subject to error, whilst fictionalism responds by retaining the usefulness of the *act* of moral discourse without accepting the *truth* of moral discourse. Abolitionists ‘try to purge thought and language of those moral elements that are found to be ... erroneous’.²⁴⁸ Abolitionism comes in two forms: ‘*partial* and *complete* moral abolitionism’.²⁴⁹ Partial abolitionism holds that ‘some subset of our moral concepts is erroneous and recommend abolishing this subset’, whilst complete abolitionism holds that ‘morality in its entirety is infected by error’.²⁵⁰ In contrast, fictionalism holds that ‘although all positive, atomic moral judgements are false ... their general acceptance

²⁴⁶ Mackie (n59) 18.

²⁴⁷ Ibid 106.

²⁴⁸ Kirchin (n2) 96.

²⁴⁹ Olson (n1) 179.

²⁵⁰ Ibid 179.

allows us ... to obtain the benefits of social cooperation'.²⁵¹ A moral fictionalist response maintains that 'it would be useful to adopt a kind of thought in which moral propositions are not genuinely believed'²⁵² but we treat moral claims as '*pretence* assertions'.²⁵³ Thus we can distinguish two kinds of fictionalism depending on the role of assertion – assertive and non-assertive fictionalism. In assertive fictionalism we 'can assert and argue that some things are right and demanded of us' but we 'do not believe in such things when reflective'.²⁵⁴ In a non-assertive fictionalist context we cannot assert moral claims but that 'effective moral communication and argumentation is still possible'.²⁵⁵

Accepting moral error theory entails being – at least – a partial abolitionist because it rejects irreducible normativity and cognitivist accounts but leaves open the adoption of non-cognitivist theories. Moral error theory does not direct us to adopt a particular conception of moral claims – it indicates what kinds of moral claims must be rejected – accordingly, it cannot help resolve moral conflict beyond excluding certain claims that if true would categorically identify moral conduct. The choice between abolitionism and fictionalism seems to turn on whether the use of moral codes promotes human wellbeing, social co-operation or some other benefit (assuming that we wish to pursue those goals). If moral codes are useful we should be moral fictionalists, if not then we can be abolitionists. But there are difficulties with moral fictionalism which suggest that it cannot perform the function assigned to it – its greatest problem is the irrationality of obeying moral claims one knows to be false.

Fictionalism implies the adoption of an irrational mindset because it requires that 'one can assert a claim and not believe it' necessitating two mutually exclusive mind states.²⁵⁶ Alternatively, it proposes engaging in a system of communication which seems distinctly odd because either communicators can mistakenly be 'taken to be asserting by others', '[t]hey may slip into the old ways of thinking and really believe that there are moral demands and values' or they know each communicator is making false statements all of which would defeat fictionalism.²⁵⁷ Fictionalism is self-defeating because of the psychological implausibility of being able to consistently

²⁵¹ Miller (n6) 113.

²⁵² Olson (n1) 181.

²⁵³ Ibid 182.

²⁵⁴ Kirchin (n2) 94.

²⁵⁵ Ibid 94.

²⁵⁶ Ibid 94.

²⁵⁷ Ibid 95.

maintain the assertive or non-assertive communicative force of moral claims whilst believing them to be false. The psychological conflict arise from accepting moral claims and submitting to their prescriptions when we know them to be false – but if they are false there is no reason to follow them. If moral fictionalists are right that moral claims promote social co-operation this not because they are *moral* claims but due to some other reason. For example, they may enhance the ability to predict how others will behave but then it is the predictability of behaviour that we want – a false moral claim is not needed to explain or justify this constraint.

For these reasons we should reject moral fictionalism. Abolitionism remains available to us but the choice between partial and complete abolition is difficult. How the term *moral concept* is constructed will affect what kind of abolitionists we are. If our notion of moral concepts include cognitivist, non-cognitivist and fictionalist accounts then rejecting two of these concepts will be partial abolitionism. If only cognitivist accounts count as moral concepts then moral error theory implies complete abolition. While this latter view is possible, the wider notion of moral concepts including both cognitivist and non-cognitivist accounts is preferable because both types of account seek to perform the same function – providing a set of behavioural constraints that agents should conform to but which are, in some way, independent of the agents to be constrained. Due to the inclusion of non-cognitivist accounts moral error theory only necessitates partial abolition of morality – although non-cognitivism may lead to a different role for moral claims in establishing a set of behavioural constraints. Non-cognitivism may give us valid moral claims to work with and before we can reach a conclusion about the implication of moral error theory for normative ethics we need to consider this alternate form of morality.

Non-cognitivist Emotivism

If moral claims are not representational beliefs about the world, what are they? One view suggests that moral claims express emotional disapproval or approval of attitudes, sentiments and actions – this is the core of emotivism. Alfred Ayer provides us with the following example:

If I say to someone, “You acted wrongly in stealing money”, I am not stating anything more than if I had simply said, “You stole money”. In adding that this action is wrong, I am not making any further statement

about it. I am simply evincing my moral disapproval of it. ... It merely serves to show that the expression of it is attended by certain feelings in the speaker.²⁵⁸

Moral disagreement, on this account, is ‘not a case of having contradictory beliefs, but rather a matter of having a clash of feelings’ and trying to change someone’s moral view is trying to change their feelings rather than convincing them of the falsity of their beliefs.²⁵⁹ This is why non-cognitivist and moral error theories are compatible – if moral claims are an expression of feeling they cannot be false in the same way that a cognitivist representational account can be false. Whether someone has a particular emotional attitude is an empirical question but what the attitude expresses is neither true nor false. Moral error has no purchase on non-cognitivism because it targets representational beliefs about mind-independent moral claims but non-cognitivism does not make this claim, hence the compatibility between them.

According to emotivism when someone says *stealing is wrong* they are not reporting anything but merely expressing their disapproval. Thus, ‘when I judge that murder is wrong, I am no more saying or reporting that I disapprove of something than I say or report that I am in pain when I cry “@*%\$!” after standing on a nail’.²⁶⁰ Emotivism uses the distinction between surface grammar – the statement of something – and deep grammar – the meaning the statement conveys. If I said *I’ve just stood on a nail!* the surface grammar would be descriptive but the deep grammar of this statement would be no more than an expression of @*%\$!. (One difficulty here is that the tone of voice also affects the meaning, if I say *I stood on a nail* I might be only describing a past event – it seems that the emotivist account requires that moral statements are always active statements about a present emotional state or attitude.) On the emotivist account the moral statement *murder is wrong* has descriptive surface grammar but the deep grammar – the meaning of the statement – is simply *Boo! Murder!*.

This emotivist conception of morality, however, is subject to the critique of implied error, which follows from the claim that moral statements are made as though wrongness or rightness is part of the factual furniture of the world. For ‘when we

²⁵⁸ Ayer (n25) 107.

²⁵⁹ Miller (n6) 25.

²⁶⁰ Ibid 34.

judge that murder is wrong, we are treating “wrong” as if it is a predicate akin to the non-evaluative predicates of our language. In other words we view *wrongness* as a property of the act of murder’.²⁶¹ Even though emotivism claims that we actually are expressing an emotion and although the surface grammar seems otherwise, we actually are referring to mind-independent properties of the world which do not exist. The emotivist can counter that either linguistically we speak of moral claims as though we are referring to a thing out-there in the world or that we ‘*project our sentiments or emotions on to the world*’ – wrongness ‘is something we *project* on to the world when we form an attitude or sentiment towards it’.²⁶²

Doing so, however, leads away from emotivism to projectivism – moral claims are not expressions of emotions rather they are properties we *add* to the world through our attitude. Thus emotivism becomes projectivism – ‘[p]rojecting is what Hume referred to when he talks of “gilding and staining all natural objects with the colours borrowed from internal sentiment”, or of the mind “spreading itself on the world”’.²⁶³ Projectivism, however, needs to explain ‘how this projection can be anything other than a *mistake* or an *error*’.²⁶⁴ That is to say, ‘[i]f we speak and think as if there is a property of goodness, although there actually is not, why isn’t our speaking and thinking in that way simply *flawed*?’²⁶⁵ A further difficulty that arises from the implied error argument is the inability to distinguish *moral* feelings from other *non-moral* feelings.

Emotivism cannot argue that moral claims “express irreducible, *sui generis*, unanalysable *ethical* feelings’ because this would result in the following circularity – a ‘[w]hat are moral judgements? Those which express ethical feelings. What are ethical feelings? Those expressed by moral judgements’.²⁶⁶ Even if we can differentiate moral expressions from non-moral expressions, we would still be unable to differentiate between different moral sentiments, because ‘bare approval does not seem to be fine-grained enough to distinguish moral goodness from moral kindness’.²⁶⁷ I agree with Crispin Wright when he states: it ‘seems to me very moot whether there is ... any distinctive mode of *moral* emotional concern, identified

²⁶¹ Ibid 35.

²⁶² Ibid 35.

²⁶³ Simon Blackburn, *Spreading the Word* (Oxford University Press 1984) 171.

²⁶⁴ Miller (n6) 36.

²⁶⁵ Ibid 36.

²⁶⁶ Ibid 40.

²⁶⁷ Kirchin (n2) 120.

purely phenomenologically and distinguished from what we feel for other kinds of values'.²⁶⁸ The alternative of explaining moral feeling as grounded in the deliberation of moral judgement is precluded because the emotivist 'wants to explain moral judgement in terms of moral feeling and not vice versa'.²⁶⁹

Yet another problem is the avoidance of extreme relativism – if non-cognitivist projectivism is the process of staining the world with our sentiments then moral claims become *dependent* on the sentiments being expressed *by each individual*. It is not as if we have collective perception or collective mental states, so we cannot be sure that the sentiments of distinct individuals are referring to the same thing or adding the same claim to the world. The risk is that non-cognitivism implies that when our sentiments change so to do right and wrong leading to unstable and highly relativistic moral claims. Miller suggests that emotivism operates 'in much the same way that changing what is on the overhead projector changes what is "on" the white wall'.²⁷⁰ It seems that non-cognitivism is not an alternative morality because it is unable to distinguish between moral and non-moral attitudes and ends in extreme relativism. If non-cognitivism fails to provide an alternative notion of morality we must conclude that there are *no moral properties and no moral claims at all*.

Non-cognitivist Quasi-realism

Simon Blackburn's more recent non-cognitivist Quasi-realist theory attempts to address these problems. Quasi-realism is 'the project of explaining how we can legitimately talk *as if* we were entitled to the assumption that there is a distinctively moral reality, even though we are not'.²⁷¹ Thus quasi-realism 'seeks to explain, and justify, the realistic-seeming nature of our talk of [moral] evaluations'.²⁷² It tries to solve this by taking us to express an attitude about moral sensibility – that is we express an 'attitude of approval towards moral sensibilities which combine [for example] disapproval of murder [or some other wrong] with disapproval of getting [others] ... to murder'.²⁷³ This attitudinal expression about moral sensibilities explains why a quasi-realist can treat propositional or cognitive moral claims as though they

²⁶⁸ Crispin Wright, 'Moral values, projection, and secondary qualities' (1988) 62 *Proceedings of the Aristotelian Society* Supplement 11.

²⁶⁹ Miller (n6) 41.

²⁷⁰ *Ibid* 39.

²⁷¹ *Ibid* 48.

²⁷² Blackburn (n130) 180.

²⁷³ Miller (n6) 55.

are real – second-order moral sensibilities determine acceptable moral attitudes towards action. Blackburn argues that the ‘use of indirect [*n*th-order attitude] contexts does not prove that an expressive theory of morality is wrong; it merely proves us to have adopted a form of expression adequate to our needs’.²⁷⁴

Quasi-realism treats our approval or disapproval of attitudes towards acts as the basis of a non-cognitivist morality, rather than treating morals as an expression of approval or disapproval of specific events. If successful this would help explain the stability of moral attitudes over time, whereas emotivism seems to suggest a *present* and *active* moral expression. Hence, in the case of murder we do not express approval or disapproval towards a particular act of killing rather we express approval or disapproval towards attitudes about killing in general. Moreover, quasi-realist projection – of higher-order moral sensibility towards attitudes – responds to natural properties thus our moral sensibilities are dependent on those natural facts. If this works then higher-order moral attitudes impose a *non-individualistic* moral response to natural facts through the process of projection. Blackburn states:

The way in which we gild or stain the world with the colours borrowed from internal sentiment gives our creations [moral claims] its own life, and so its own dependence on facts.²⁷⁵

The suggestion is that projection of moral sensibilities can establish a link between moral claims and natural properties which are sustained independently of our individual perspective even ‘were our sentiments to alter or disappear’.²⁷⁶ The difficulty for quasi-realism is that the non-cognitivist element requires ‘that our conative attitudes cannot be correct or mistaken’ but the quasi-realist element (the aim of talking as if moral claims are correct or mistaken) requires that ‘these attitudes *can* be correct or mistaken’.²⁷⁷ It is not clear how moral claims can gain a life of their own from higher-order attitudes, for if our higher-order attitudes change then our moral claims *would* also change – the problem is simply replicated at a higher-order level. Similarly, Thomas Scanlon is critical of second-order desires over first-desires

²⁷⁴ Blackburn (n130) 195.

²⁷⁵ Ibid 219, fn.21.

²⁷⁶ Ibid 219, fn.21.

²⁷⁷ Derek Parfit, *On What Matters*, vol 2 (Oxford University Press 2011) 400.

because ‘they are just two desires that conflict with each other’.²⁷⁸ It seems that a similar question can be asked of quasi-realism – how can first-order moral claims that are dependent on the projection of second-order moral sentiments remain fixed even when our moral sentiments change? Moreover, how do we know which moral sentiments are correct? How do we know if a moral sentiment expressing approval towards duelling is wrong?

We can see that the weakness of non-cognitivist theories of morality boils down to two main concerns. First, non-cognitivism seems unable to explain how moral claims are anything more than individualistic attitudes. For this not to be the case non-cognitivism would need to entail that distinct agents can possess the same mental state towards an action but while people may have *similar* attitudes they cannot share mental states to the degree necessary to ensure that moral claims *mean the same thing* to different people. Second, it cannot adjudicate between conflicting moral sentiments, expressions, projections or attitudes because non-cognitivist moral claims are neither true nor false and cannot be measured on a true-false scale. Nor can it distinguish between moral and non-moral sentiments and attitudes. Even if these problems were resolved we would not have a definitive list of moral claims but a situation of multiple valid moral claims based upon different attitudinal responses. On the other hand, non-cognitivism is more easily able to explain the presence of normative force because attitudes contain a component of approval or disapproval that necessarily implies we have both reason and motivation to act according to that attitude. Although non-cognitivism cannot account for claims of irreducible normativity, it can explain reducible normative claims because attitudes express how we want the world to be.

If moral error theory about cognitivist morality is correct and non-cognitivist moral claims are the only claims available to us, we *must accept* the following conclusion – a system of irreducibly normative moral claims is impossible because such claims are false and non-cognitivism (if successful) results in multiple valid moral claims. Consequently, we must conclude that first-order normative systems cannot be founded on moral claims because no moral claim applies to all agents. But this conclusion challenges not only the basic understanding of morality as a *shared practice* it also suggests that individuals are *unknowable* to each other – is there really

²⁷⁸ Thomas Scanlon, *What We Owe Each Other* (Harvard University Press 2000) 55.

no other alternative theory which can explain moral claims as a shared understanding of behaviour? The last bastion of a common morality seems to be language and the notion that shared linguistic systems confer meaning inter-subjectively between individuals – perhaps inter-subjective discourse morality can resolve the relativism of the other non-cognitivist theories whilst also explaining the source of moral claims.

Morality as Discourse

The idea of morality as discourse sources the validity of moral statements in the language and concepts used within communities. On this view morality is akin to a language-game – ‘the propriety of a moral term is ensured primarily by linguistic rather than moral values’.²⁷⁹ In contrast, traditional conceptions of moral claims ‘presume a prioritisation of extension over meaning’.²⁸⁰ Discourse based morality take a reversed view – *meaning has priority over extension*. Moral claims do not refer to facts or attitudes rather linguistic morality prioritises meaning as constituted by the role moral claims play in language. On this account, the moral claim *homosexuality is unacceptable* or *homosexuality is acceptable* does not refer to facts about the world or to the attitudes of agents. These terms acquire their meaning from the way speakers of a community use language (how they use the word acceptable or unacceptable assigns positive and negative connotations) and normativity is built into the languages in use.

We will get moral disapproval at the end of the process not because murder [or some other moral term] violates some universal and rationally available standard of morality, but simply because of the meaning of the word “murder”.²⁸¹

Essentially, a moral word brings ‘with it a normative tone’ and ‘[a]s long as we understand what a word means, its normative value is apparently waiting for us’.²⁸²

Yet as the example of homosexuality highlights, multiple discourses can be present at the same time within a single community but because discourse grounds moral propositions. ‘[A]ny statements we might make or beliefs that we might have,

²⁷⁹ Brassington (n3) 129.

²⁸⁰ Ibid 129.

²⁸¹ Ibid 131.

²⁸² Ibid 131.

though formative of ... [agents], are “incomplete”²⁸³. Any agent who makes a moral statement must recognise the limits of himself and the world in which he finds himself and one of those limits is the *discursive limits* of moral claims. Thus ‘[o]ne’s moral “knowledge”, whether that be individually or culturally ordained, is *only* a perspective’.²⁸⁴ Acknowledging this ‘gives us pause in our moral certainties’ because we accept that moral propositions are limited to the discourse in which they occur.²⁸⁵ Once again we encounter the territorial limits of a theory leading towards relativism – albeit a variety of relativism tied to linguistic communities rather than individuals. Even if we accept that a discursive account of morality exists, it cannot provide a definitive account nor can it adjudicate between different linguistic communities (whether intra- or inter-socially).

Normative constraint, therefore, still requires a non-moral justification because discursive-morality is *incomplete*. It remains open at this point whether Jürgen Habermas’s theory of communicative action can address this problem – unfortunately, I possess insufficient knowledge of Habermas’s theory to suggest an answer.²⁸⁶ Nevertheless, discursive morality, like non-cognitivism, is unable to justify constraint for all agents, provide a means of distinguishing between correct and false moral claims or avoid relativism. Thus, the conclusion that cognitivist morality is false, that non-cognitivism cannot provide non-relativistic moral claims and that we must reject moral claims as founding a system of first-order normative behavioural constraints applicable to all agents, is valid and must be accepted.

Conclusion: Post-Moral Error Theory and Shaping Normative Ethics

Moral error theory demonstrates that objective, prescriptive, mind-independent moral claims are not truth-apt because irreducible normativity is a primitive notion which can only be asserted. Consequently, agents cannot rely on the discovery of moral claims that are binding on all agents to solve interaction problems. Moral claims are relativistic (either to individuals or social units) and as such they cannot function as

²⁸³ Ibid 166: I understand ‘Dasein’ to be roughly equivalent to agent although the important difference from ‘traditional’ notions of agents is that an agent is not separable from the world – agents are constituted by the world in which they find themselves.

²⁸⁴ Ibid 167.

²⁸⁵ Ibid 173.

²⁸⁶ See: Jürgen Habermas, *The Theory of Communicative Action: Reason and the Rationalisation of Society* vol 1 (Polity Press 2004); Jürgen Habermas, *The Theory of Communicative Action: Lifeworld and Systems, a Critique of Functionalist Reason* vol 2 (Polity Press 2004).

the basis for a system of general constraint for a heterogeneous population. In a heterogeneous population there will be disagreement between agents concerning justifiable constraint but moral error theory demonstrates that different moral claims cannot be true because all moral claims are either uniformly false or not truth-apt. Even if every member of society accepted that there are irreducibly normative reasons and agreed on what irreducibly normative moral claims there are, disagreement would still remain valid because individual agents can reject such principles on the basis of moral error theory (although the practical problems would be managed by enforcement mechanisms as all agents would agree to the same set of moral rules). There are, broadly speaking, three responses to the crisis provoked by moral error theory. First, we embrace moral nihilism (referred to as abolitionism earlier) and reject all moral claims from inclusion in the debate about justifiable constraint. Second, we might conclude that we should retain moral claims because ‘[m]oral talk can help coordinate desires and interests more readily than other methods’.²⁸⁷ Third, we might conclude that we should construct a system that provides a non-moral procedure for generating normative constraint between agents who possess different metaethical and normative stances.

Moral nihilism (or complete moral abolitionism) would reject any and all moral claims. We might take this view because we believe, as Mackie seems to have done, that moral claims are only really *moral* if they have an irreducibly normative favouring relation as a component. If moral claims are reducible to some other factor – say the desires, goals, rationality or attitudes of agents – they would cease to be distinctly *moral* claims. Any moral claims made which lack the property of irreducible normativity would be more akin to conventions or standards of conduct rather than morality. Thus if moral claims lost the distinctive prescriptive force that irreducible normativity provides conceptions of right and wrong, justice and injustice, good and evil would become irrelevant and moral claims would cease to have any justificatory force for constraint. But without moral constraint, so the argument might go, a major source of regulating personal and inter-personal conduct would be lost to us. Yet if moral claims are false or lack the normativity to bind disparate agents their use to justify constraint would be mere contrivance, a blight, a curse, an unjust malady hampering the resolution of conflict between agents by inducing them to

²⁸⁷ Kirchin (n2) 93.

believe their moral claims are absolute, correct, objective and normatively authorised as the *right way* to conduct oneself. If anything history has shown how much misery, suffering and conflict this approach actually generates because it increases and encourages fanaticism and a conviction in the absolute truth of a particular perspective. Axel Hägerström suggests that the absence of irreducibly normative moral claims may help prevent fanaticism in applying moral beliefs. Hägerström argues that ‘the belief that there is a unique moral truth that one has come by tends to lead to fanaticism’, while moral error theory undermines belief in a unique moral truth and thus ‘promote[s] toleration and understanding of other people’s beliefs and customs’.²⁸⁸

For these reasons the second option of retaining moral discourse in its current form is unattractive for it would simply see the continued use of moral propositions. It has been suggested that the use of current moral discourse is justified because ‘the regulative and coordinating functions they facilitate are of such vital importance to us’.²⁸⁹ However, a problem arises similar to the one we saw with fictionalism, namely the requirement of possessing two conflicting mental states. We would have to use a system that we know is flawed or in error, consequently it is not clear how we could accept moral propositions as normative nor is it clear how this could justify imposing constraint on agents who disagreed. Recognising this, Olson suggests that we adopt the position of Moral Conservationism in which we use ‘moral thinking and reasoning when we find ourselves in “morally engaged” everyday contexts and that we turn to [metaethical] ... thinking and reasoning only in “detached and critical contexts”’.²⁹⁰ In other words, ‘the more reflective beliefs [in moral error theory] are suppressed or not attended to’ in everyday moral situations.²⁹¹ This is not a solution, not only because anyone who wanted to reject the applicability of a moral claim would be able to engage in metaethical reasoning – they could bring reflective beliefs into the everyday to undermine the moral proposition(s) – but also because the very problem metaethics should resolve is the conflict between competing everyday moral claims. Adopting Moral Conservationism is a betrayal of the metaethical project and conclusions that moral error theory presents.

²⁸⁸ Olson (n1) 56-7, fn.43.

²⁸⁹ Ibid 143.

²⁹⁰ Ibid 192.

²⁹¹ Ibid 192.

If moral claims cannot provide a unique set of normative prescriptions an alternative method of developing normative constraint on inter-personal conduct is needed but crucially any such system must operate on the basis that there are no definitive moral stances – an approach that is compatible with (although not necessitating) moral nihilism. Mackie usefully distinguishes between ‘morality in the broad sense [which] would be a general all-inclusive theory of conduct’ and the ‘narrow sense, [where] morality is a system of a particular sort of constraint on conduct’.²⁹² The aim of morality in this narrow sense is to ‘protect the interests of persons other than the agent and which present themselves ... as checks on his natural inclinations ... to act’.²⁹³ The narrow sense of morality as a system of constraints between agents is our concern here and it is the *only* sense of morality that can justify constraint for all agents *without* agreement upon the fundamental nature of moral claims. Agents seeking to fulfil their preferences²⁹⁴ and resolve interaction problems can best achieve this through a system of constraint but because moral claims are in error they are a hindrance to constraint rather than a source of strength – except in the psychological sense of inducing guilt and regret in agents – because they create political and moral deadlock and fanaticism.

Constraint – as a system aimed at promoting the preferences of agents – can be grounded on *reducible normativity* because this form of normativity is reducible ‘to facts about what promotes desire satisfaction’²⁹⁵. This reframes normative constraint as a system based upon the characteristics of agents themselves, rather than the traditional conception of morality as a set of external prescriptions that agents must obey regardless of their particular agenda. Reducible normativity is the only possible basis for constraint because it ties constraints to achieving an agent’s purpose giving them a reason to comply with those constraints – a reason now absent after the coup de grace of moral error theory on cognitivist morality. This alternative is needed because ‘[n]othing has altered or will alter the importance of being able to make and keep and rely on others keeping agreements’.²⁹⁶ The question, then, becomes what constraints or rules should agents agree to? This is the focus of the next chapter.

²⁹² Mackie (n59) 106.

²⁹³ Ibid 106.

²⁹⁴ Here, preferences is a catch all term for the outcomes, goals, objectives and circumstances that agents want to realise in the world.

²⁹⁵ Olson (n1) 121.

²⁹⁶ Mackie (n59) 123.

Chapter 2

Can Rational Constraint Secure Co-operation?

Introduction

‘Morality faces a foundational crisis. Contractarianism offers the only plausible resolution of this crisis’.¹ This crisis arises because ‘moral language fits a world view that we have abandoned – a view of the world as purposively ordered’ by right and wrong, justice and injustice, good and evil.² Morality is taken to be one foundational claim or a set of claims by which all questions concerning human action and interaction can be answered. For such claims to achieve this they must be ‘intrinsically prescriptive entities’ – moral truths would need to have ‘to-be-pursuedness somehow built into it’ but such moral claims, as I argued in the last chapter, are metaphysically queer.³ Yet, the ‘main tradition of European moral philosophy includes the contrary claim’ because it holds that moral claims provide a conclusive reason to act in accordance with moral prescriptions.⁴ As the previous chapter demonstrated, the absence of objectively prescriptive moral claims means that morality fails ‘to provide a justificatory framework for moral behaviour and principles’.⁵ It fails to function as ‘a system of a particular sort of constraints on conduct – one whose central task is to protect the interests of persons’ but this function must be performed, otherwise agents have no guidance on how to collectively regulate their conduct towards each other to maximise their preference fulfilment.⁶

Hobbesian contractarianism *can* perform this function without relying on moral propositions because its methodology is based upon rational self-interest – in this way contractarianism is the solution to the foundational crisis of morality. Furthermore, it does not require a moral claim that ‘ethical conduct is acceptable from a point of view that is somehow *universal*’⁷ or ‘from the point of view of the

¹ David Gauthier, ‘Why Contractarianism?’ in Peter Vallentyne (ed) *Contractarianism and Rational Choice* (Cambridge University Press 1991) 15.

² Gauthier, ‘Why Contractarianism?’ 16.

³ John Mackie, *Ethics: Inventing Right and Wrong* (Penguin Books 1990) 40.

⁴ Ibid 30.

⁵ David Gauthier, *Morals by Agreement* (Oxford University Press 2001) 2.

⁶ Mackie 106: If individuals separately regulated their conduct to each other this would not provide the consistency and applicability needed for a *system* of constraint for all agents.

⁷ Peter Singer, *Practical Ethics* (Cambridge University Press 2011) 10.

universe'.⁸ Nor does contractarianism need Thomas Scanlon's idea of mutual recognition⁹ as a fundamental moral premise that means we are 'unreasonable ... [in] refusing to take other people's interests into account'.¹⁰ All three notions of a moral point of view justify constraint by requiring that agents accept the interests of others as a reason for or against certain actions. But we cannot assume that the interests of others are relevant to constraint for that would be to use a moral premise and thus would not solve the foundational crisis.

Practical rationality does provide a solution because it takes an agent and their goals as given, works out what the agent needs to do to achieve those goals and – where co-operation is required – tells agents to comply with constraints on interaction with other agents. These constraints are justified because these 'are constraints that an agent imposes on himself for the sake of some expected benefit to himself'.¹¹ Rational constraint is normative constraint because these constraints are reducibly normative – they 'are reducible to facts about what promotes desire satisfaction', they are reducible to facts about agents and their environment.¹² Thus, the basic premise of contractarian constraint is the normative claim that an agent *should* co-operate with other agents because this is the best means of achieving an agent's goals – this serves 'both for prescription and for critical assessment' of constraints.¹³ This constraint-benefit rationale is the core around which the analyses of human rights law (in chapter three) and the welfare of the child test (in chapter four) are built.

In what follows, I analyse David Gauthier's contemporary neo-Hobbesian theory showing that, while self-regulation can take some distance, second-order constraint is needed bring about the most optimal (that is preference maximising) co-operative outcomes. Considering Hobbes's all powerful sovereign and Gauthier's self-regulating co-operative agents will illuminate the difficulties in ensuring compliance amongst agents and – most importantly – highlight the role of legal rules as second-order constraints to bring about more efficient co-operative outcomes. The increasing heterogeneity of agents in populations subject to constraint means that a

⁸ Henry Sidgwick, *Methods of Ethics* (7th edn, University of Chicago Press 1907) 382.

⁹ Mutual recognition is based on "our relations with others, as rational creatures who recognise many of the same reasons and can recognise the value of each other" in Thomas Scanlon, *What We Owe Each Other* (Harvard University Press 2000) 77.

¹⁰ Thomas Scanlon, *What We Owe Each Other* (Harvard University Press 2000) 33.

¹¹ Jon Elster, *Ulysses Unbound* (Cambridge University Press 2000) 4.

¹² Jonas Olson, *Moral Error Theory* (Oxford University Press 2014) 121.

¹³ Gauthier, *Morals by Agreement* (n5) 3.

single authoritarian sovereign is untenable as a means of forcing compliance, while Gauthier's self-regulating agents should lead to exploitation when this is the most rational course of action for an agent. The solution to ensuring compliance is the use of second-order constraints to modify payoffs which provides the best means for maintaining co-operation and preventing the gradual dissolution of co-operation through increasing exploitation. For example, an all powerful sovereign cannot ensure that every agent pays their taxes and for each individual agent it is rational to evade the taxation system as much as possible while trying to ensure all other agents contribute through taxation. To ensure the greatest degree of compliance with taxation second-order constraints can increase the costs of tax evasion so that it is rational for each individual agent to pay their taxes. The analysis of the contractarian approach leads to the following conclusion: the most efficient system of constraint uses second-order constraints to modify the payoff of interactions to increase the frequency of co-operative outcomes. This conception of second-order constraint leads to a contractarian theory of judicial regulation wherein justified legal rules are those that are rational constraints.

Constraint through Coercion

Constraint makes possible co-operation which is beneficial because without co-operative schemes we face the spectre of unrelenting competition. The theoretical state of nature, of 'every man, against every man' where no agent is constrained and they are free to do as they wish even though this leads to the worst outcome for all agents.¹⁴ This state arises due to 'three principal causes of quarrel ... [f]irst, competition; secondly, diffidence [distrust]; thirdly, glory. The first, maketh men invade for gain; the second, for safety; and the third, for reputation'.¹⁵ However, because agents are self-preserving they are led to seek peace (co-operation) because of the 'desire of such things as are necessary to commodious living ... [a]nd reason suggesteth convenient articles of peace, upon which men may be drawn to agreement'.¹⁶

Hobbes proposed two fundamental articles of peace. The first is 'that every man, ought to endeavour peace, as far as he has hope of obtaining it; and when he

¹⁴ Thomas Hobbes, *Leviathan* (first published 1651, Oxford University Press 1996) 84.

¹⁵ *Ibid* 83.

¹⁶ *Ibid* 86.

cannot obtain it, that he may seek, and use, all helps, and advantages of war'.¹⁷ Peace is worth pursuing because 'men can be enabled to maintain themselves without facing the competition of their fellows, then the basic cause of hostility among men will be removed'.¹⁸ A modern expression of this notion would be that if agents are in competition with each other they will invest resources in protecting themselves rather than investing in other more productive pursuits. Rational self-interest leads people to seek a stable secure system of peace in which to live. Where constraint and co-operation breakdown, those who seek peace must abandon their efforts and resort to the unrestrained conduct of war.

The greatest obstacle to co-operation is ensuring that multiple individuals will comply with the constraint that secures co-operation. To solve this Hobbes proposed the second law of nature which states that individuals 'lay down this right to all things; and be contented with so much liberty against other man, as he would allow other men against himself'.¹⁹ But rather than simply setting aside their right to war, agents transfer their right to the all-powerful Hobbesian sovereign who 'tie[s] them by fear of punishment to the performance of their covenants' – the sovereign being one individual who controls the strength of the multitude to punish those who violate or break the covenant.²⁰ The Hobbesian sovereign is necessary because compliance is 'contrary to our natural passions, that carry us to partiality, pride, revenge, and the like'.²¹ 'Human beings are not, as a group, rational enough to be able to institute moral conventions, and hence must create a sovereign who can use his power to generate them'.²²

The mutual laying down of the unlimited right of nature is the means by which constraint is introduced through the creation of an enforcement mechanism. In Hobbes's original political theory this mechanism is a monarch but the development of contemporary society has rendered the sovereign an inadequate enforcement mechanism because of the large heterogeneous populations that exists. A far more efficient, cost effective and beneficial system would have agents voluntarily act

¹⁷ Ibid 87.

¹⁸ David Gauthier, *The Logic of Leviathan* (Clarendon Press 2011) 19.

¹⁹ Hobbes (n14) 87.

²⁰ Ibid 111.

²¹ Ibid 111. Also see Elster who discusses pre-commitment devices to restrain irrational behaviour due to the passions in his *Ulysses Unbound* (n11).

²² Jean Hampton 'Two Faces of Contractarian Thought' in Peter Vallentyne (ed) *Contractarianism and Rational Choice* (Cambridge University Press 1991) 34.

according to the system of constraint (think of the resources needed to monitor every agent in a contemporary population, let alone the cost of employing enforcement officials) – but without the fear of a sovereign, how can we ensure compliance? David Gauthier in his *Morals by Agreement* seeks to develop Hobbes’s theory of constraint into one with which agents comply even in the absence of a threatened punishment and even when they could do better by betraying the constraint system. He aims to show that human agents *are* sufficiently rational to institute constraints and comply with them without the threat of force to ensure compliance.

Instrumental Rationality and Constraint

Gauthier ‘defends an instrumental conception of practical rationality, according to which a choice is rational if and only if relative to the agent’s beliefs it is the most effective means for achieving the agent’s goals’.²³ As Gauthier states:

We shall argue that the rational principles for making choices, or decisions among possible actions, include some that constrain the actor pursuing his own interest in an impartial way. These we identify as moral principles.²⁴

Using practical rationality we can ‘suppose that some moral principles may be understood as representing joint strategies prescribed to each person as part of the ongoing co-operative arrangements that constitute society’.²⁵ Or more clearly, by co-ordinating the actions of multiple agents the behaviour of each agent is regulated in such a way that co-operation is achieved and maintained. Moral claims, therefore, are dependent upon being chosen through rational choice that covers both the selection of constraints and compliance with those constraints. Note that on this conception *moral constraints are rational constraints* but this terminology is confusing because it suggests that constraint is justified by something additional to rational choice – for Gauthier the structural impartiality of rational constraint applied to an entire population of agents is what makes these constraints *moral*. However, what is actually doing the theoretical and explanatory work is *rationality*. Due to the rejection of

²³ Peter Vallentyne, ‘Gauthier’s Three Projects’ in Peter Vallentyne (ed), *Contractarianism and Rational Choice* (Cambridge University Press 1991) 5.

²⁴ Gauthier, *Morals by Agreement* (n5) 3.

²⁵ *Ibid* 168.

moral claims as justifying constraint, constraint should be taken to mean *rationaly justified constraint* because this more accurately reflects the theoretical concepts underpinning constraint and the method by which we decide on constraints (although references to moral constraint will be retained in quotations). Furthermore, these constraints *are* impartial in the sense that they cover all interaction between all agents. Although this does not assume the agents are morally equal, rather agents are participants in interaction and it is this that requires the consideration of all – for if some agents are excluded they cannot be held accountable to the system of constraint and they may have a reason to overturn that system.

Using instrumental rational choice theory means that there ‘are no external norms for assessing someone’s preferences ... except the formal coherence properties’ of completeness, transitivity, monotonicity and continuity.²⁶ Completeness requires that ‘any two members [of a set of preferences] be comparable on the basis of the relationship’ between preferences.²⁷ The second condition is transitivity, which holds that preferences should hold through a chain of preferences. Consider eating fruit: if someone ‘prefers eating an apple to eating a pear and eating a pear to eating a peach, then if transitivity holds he prefers eating an apple to eating a peach’.²⁸ The third condition is monotonicity which, similarly to transitivity, orders preferences of outcome-groups based upon the preferences between outcomes within those groups. The following example clarifies this statement:

Consider two lotteries, X and Y, that differ only in one respect - one of the prizes (or outcomes) in X, P, is replaced by another prize in Y, Q (with, of course, the same probability). Monotonicity requires that if Q is preferred to P, then Y is preferred to X.²⁹

The final condition is continuity which ensures that if there are ‘three possible outcomes, X, Y and Z, such that X is preferred to Y and Y to Z. Continuity requires that there is one and only one lottery, with X and Z as prizes, that is indifferent to Y’.³⁰ In other words, if an agent has the option of gambling on X or Z, or the certainty

²⁶ Vallentyne, ‘Gauthier’s Three Projects’ (n23) 6.

²⁷ Gauthier, *Morals by Agreement* (n5) 39.

²⁸ *Ibid* 40.

²⁹ *Ibid* 44.

³⁰ *Ibid* 44.

of Y he will be indifferent between gambling on X and Z or guaranteeing Y. If someone prefers – in the following order – apples, peaches and pears he will be indifferent between gambling on gaining apples or pears and gaining, for certain, a peach.

Once these formal criteria are met an agent will have considered preferences – ‘preferences are considered if and only if there is no conflict between their behavioural and attitudinal dimensions and they are stable under experience and reflection’.³¹ These conditions are the formal requirements rational choice imposes upon the preferences of agents but beyond this there are no further requirements concerning the substantive goals of those preferences. For example, prospective parents who have a preference for a disabled child cannot be prevented from holding that preference but they cannot simultaneously prefer a disabled child and a non-disabled child. The validity of preferences is derived from formal cohesion between them – they ‘are not subject to rational assessment in terms of their contents’.³² The role of rational choice ensures that contractarian theory is ‘practical and *person-centred*’³³ because the underlying ‘contractarian strategy is grounded in a subjective instrumental and relativistic conception of rational choice’.³⁴

Consequently, the content of an agent’s considered preferences is left open. As long as intrapersonal preferences are formally rational the content of them can be determined by any means – it is the choices made in pursuit of a preference that is subject to contractarian constraint because this is the relevant area of concern in interaction. Thus, when we come to discuss the role of the welfare of the child test in the regulation of reproduction, we can favour person-affecting as the basis of rational constraint because we know that this occurs but agents who hold an impersonal conception of harm cannot be made to disregard their preferences – hence, they should not be forced to adopt a person-affecting basis for *their* procreative decisions (although person-affecting harm will apply in their interaction with others). However, this also means that they cannot require that impersonalism forms the basis of rational constraint because we have no convincing reason to think impersonal harm can occur to those who do not exist. Under the theory of rational constraint, whether I have a

³¹ Ibid 32-33.

³² Ibid 38.

³³ David Copp, ‘Contractarianism and Moral Skepticism’ in Peter Vallentyne (ed) *Contractarianism and Rational Choice* (Cambridge University Press 1991) 199.

³⁴ Copp 198.

preference for murder or torture is less important than the constraints that other agents may require of me to secure co-operative interaction and whether I *act* to fulfil those preferences.

Co-operation, Constraint and the Compliance Dilemma

There are two distinct types of choice situations. First, parametric situations ‘in which the actor takes his behaviour to be the sole variable in a fixed environment’ – in parametric situations there is no need for constraint so an agent will select their most preferred outcome thus maximising their utility – which means maximising one’s preference fulfilment³⁵. In strategic interaction the behaviour of one agent will affect the behaviour of other agents which in turn will affect the first agent’s actions and so on. Consider the following example: Patrick is the only person who can eat from a buffet, if Patrick’s most preferred food is lasagne then he can take all of it if he wishes – there is no need for constraint here because it is a *parametric situation*.

However, now imagine that Patrick is standing in a queue for the buffet with others who also seek to fulfil their preferences by selecting their preferred food. In this *strategic situation* Patrick’s choice will be affected by the decisions of those before him and his decisions will affect the decisions of those behind him. Patrick prefers lasagne but if this is unavailable – because Edward has eaten all of it – he will take meatballs instead. If meatballs are the most preferred food of someone behind him, George say, then George’s preference fulfilment will be frustrated by Patrick’s choice (assuming he takes all of or the remaining meatballs). Constraint becomes relevant in these situations because these individuals are in competition for their most preferred outcome. If Patrick could agree with Edward to share the lasagne and both could agree with George to leave the meatballs for him – let us assume neither Patrick nor Edward will choose meatballs if lasagne is available – then all three would fulfil their preferences to a greater extent through co-operation than through non-cooperation because each agent would always get their most preferred meal.

Consequently, agents that interact with other agents can co-operate to maximise their preference fulfilment which they would be unable to secure in non-co-

³⁵ Gauthier, *Morals by Agreement* (n5) 21. As Prof. Søren Holm has pointed out that an agent could aim at *satisfying* their preferences rather than *maximising* them. However, in a parametric case it would be irrational to fail to maximise one’s preferences. Furthermore, in bargaining if others know you will settle at less than maximum they would have a reason to withhold the maximum benefits from you. For these reasons I will retain the maximising conception of instrumental rationality although agents may settle for less than optimal maximisation of preference fulfilment.

operative competitive interaction. But what ensures that self-interested individuals cooperate with each other? For Gauthierian contract theory, individuals should cooperate not because they are under threat from all others or a sovereign but because their self-interest calls for such co-operation. This is due to the fact that the ‘sources of satisfaction and dissatisfaction are not in fixed supply, so that by appropriate interaction overall costs may be lessened, and overall benefits increased’.³⁶ But why should Edward co-operate with the others when he can simply take all the lasagne himself? The answer is that *repeated interaction* makes it rational to comply with constraint if future interaction is contingent on an agent complying. If Edward knew that he would be queuing with Patrick and George over many occasions (perhaps because they all work for the same organisation) he may co-operate because he cannot always ensure that he will be first. Yet, Gauthier intends his theory to apply *even* in isolated one-off interactions. Thus:

[In] a purely isolated interaction, in which both parties know that how each chooses will have no bearing on how each fares in other interactions. ... A constrained maximiser chooses to co-operate if, given her estimate of whether or not her partner will choose to co-operate, her own expected utility is greater than the utility she would expect from the *non-co-operative outcome*.³⁷

This is the ‘highly significant feature’ of Gauthierian contract theory because ‘if successful, it entails [co-operation] ... when interacting with you, even if you and I will only interact on this single occasion’.³⁸ If this theory can demonstrate that constraints apply even when agents participate in one-shot interactions we will have a theory of constraint that agents will always obey. This would be a significant achievement because it would solve the compliance dilemma that arises from the following reasoning: ‘while I benefit from others’ impartial constraint, my own constraint is a pure cost to me’.³⁹ That is, in situations when Edward can more effectively maximise his preferences by convincing others (Patrick and George) to

³⁶ Gauthier, *Morals by Agreement* (n5) 114.

³⁷ *Ibid* 170 (my emphasis).

³⁸ Holly Smith, ‘Deriving Morality from Rationality’ in Peter Vallentyne (ed) *Contractarianism and Rational Choice* (Cambridge University Press 1991) 235.

³⁹ Peter Danielson, ‘Closing the Compliance Dilemma’ in Peter Vallentyne (ed) *Contractarianism and Rational Choice* (Cambridge University Press 1991) 292.

constrain their behaviour whilst not constraining Edward's that is the most rational action to take. But rational agents know that everyone will think this way leading mutual defection even if they agree to constraint making co-operation impossible.

The challenge we face is as follows: agreeing to constraint is rational but when the time comes to comply with that constraint, the rational act may be to defect from or fail to comply with it when that gives the agent a higher payoff than compliance. If it can be shown that rational choice theory requires the acceptance and compliance of constraints regardless of the benefits of defection, then, 'there is a rational solution to the problem of compliance ... [and] it follows that no enforcement mechanism (that imposes sanctions on those who do not comply) is needed' for no defection will occur.⁴⁰ Solving the compliance dilemma is the major focus of this chapter and, after showing that self-regulation amongst agents fails to ensure the optimal outcome, the compliance dilemma provides the justification for second-order constraints in the form of judicial regimes and their rules.

Note that we have two interpretations of the compliance dilemma: on Hobbes's account agents break their covenants because of the passions which move them towards counterproductive conduct (pride, revenge, greed etc.) while on rational constraint account the compliance dilemma arises in those cases where the largest pay off comes from defecting from constraint. Hobbes's account is similar to Jon Elster's self-discipline theory in which agents accept constraint (in the form of pre-commitment devices) to protect themselves from time inconsistency and irrationality.⁴¹ Both Hobbes and Elster propose pre-commitment devices to limit the impact of the passions, time inconsistency and irrationality although they differ in that Hobbes was concerned with interpersonal conduct while Elster is concerned with intrapersonal conduct.⁴² Rational constraint in contrast, seeks to produce an internal system for compliance rather than Hobbes's external coercive sovereign or Elster's pre-commitment devices.

However, the character Maximilian Bercovicz (from Sergio Leone's 1984 epic *Once Upon a Time in America*) highlights the significant problem the compliance dilemma presents. Bercovicz betrays his friends, kills two of them, fakes his own death and steals the million dollars they have made in order to reinvent himself, going

⁴⁰ Vallentyne (n23) 10.

⁴¹ See: Elster, *Ulysses Unbound* (n11).

⁴² See: Elster, *Ulysses Unbound* (n11) Chapter 2 for Elster's discussion on constitutions as pre-commitment devices.

on to become a successful politician. Here, the constraint of loyalty to one's friends fails because the payoff Bercovicz receives for defection is sufficiently high that ending all interaction with his friends does not result in a net loss for him. On the contractarian theory of rational constraint we *must* say that Bercovicz's betrayal is rationally permitted (in fact that it is rationally prescribed) but this outcome is clearly very bad for his friends (they have either been killed or left destitute). If they had been able to anticipate Bercovicz's actions they would have taken steps to prevent him and protect themselves. It is at this point that Hobbes's solution – the coercive power of a sovereign – arrives on the scene. Hobbes argues that 'nothing is more easily broken than a man's word' and 'bonds ... have their strength ... from fear of some evil consequence upon rupture'.⁴³ Thus a sovereign who can command the power of all men is needed to coercively enforce compliance with the covenant and to punish those who break it. In this case the sovereign could impose a penalty (say imprisonment for life or execution) on Bercovicz such that he is made *worse off* by defection than by complying with constraints – for '[t]o agree to act at all is to agree that complaint is in order for nonperformance' justifying the imposition of sanctions.⁴⁴

This is exactly the kind of case we expect moral constraint to prevent but we cannot rely on moral constraint as a consequence of moral error theory. Hobbes responds to this problem by introducing a pre-commitment and enforcement mechanism (the sovereign) while Gauthier wants to show that it is *rational* to comply with agreements *even if* this leads to a lower payoff than defection. For Hobbes, the compliance dilemma arises because of the passions of agents while for the theory of rational constraint (and for Gauthier), the problem arises due to the rational maximising approach which will sometimes require defection *but* when this is known it will lead to worse outcomes because no one will co-operate.

The Dispositional Solution to the Compliance Problem

To solve the compliance dilemma, Gauthier introduces the concept of rational dispositions under which one adopts the disposition that affords the greatest utility

⁴³ Hobbes (n14) 88.

⁴⁴ Jan Narveson, 'Contractarian Rights' in R.G. Frey (ed) *Utility and Rights* (University of Minnesota 1984) 166.

(fulfilment of preferences). But what is meant by *disposition* and why does Gauthier see dispositions as the solution?

First, an agent ‘benefits from her disposition, not in the choices she makes, but in her opportunities to choose’.⁴⁵ Second, as co-operation relies on expectations concerning the actions of others, dispositions provide us with the necessary information for us to form these expectations (assuming the agents are transparent – an issue which is discussed later). ‘The essential point ... is that one’s disposition to choose affects the situations in which one may expect to find oneself’ – that is, an agent creates more opportunities for interaction by accepting constraints and complying with co-operative outcomes that are not directly maximising.⁴⁶ They also convey information to those with whom we interact telling them that we *will* behave in certain ways because ‘dispositions may be the most effective way of compensating for the weakness’ of agents, implying that dispositions cannot be easily changed once taken up thus other agents can rely on consistency in our behaviour.⁴⁷

Once other agents know of our disposition then they will interact with us to a greater extent than they otherwise would, thus increasing the overall benefits of co-operation because although we receive a lower payoff per interaction we engage in more interaction than we could by acting otherwise. Dispositions are, therefore, the solution to the compliance dilemma because they allow us to predict profitable interaction with another agent. The problem of justifying morality becomes one of providing a self-interested rationale not directly for behaving morally but for being moral, for having the disposition to reason morally’.⁴⁸ For this to be successful agents must publicise their disposition to co-operate. However, deceptive non-co-operative exploitative agents will mimic this publication to fool others into co-operating with them. If this deceptive behaviour becomes widespread enough no agents would be able to rely on the publicised information about an agent’s behaviour (we shall look at the problem of deception later on).

Dispositions are ‘Gauthier’s ingenious solution [which] involves moving from the appraisal of one’s choices to the appraisal of one’s dispositions to choose’.⁴⁹ As a result of this dispositional approach, Gauthier proposes the constrained maximiser

⁴⁵ Gauthier, *Morals by Agreement* (n5) 183.

⁴⁶ *Ibid* 183.

⁴⁷ *Ibid* 185-6.

⁴⁸ Geoffrey Sayre-McCord ‘Deception and Reasons to be Moral’ in Peter Vallentyne (ed) *Contractarianism and Rational Choice* (Cambridge University Press 1991) 184-5.

⁴⁹ Copp (n33) 200.

(CM) disposition as the uniquely rational solution to the compliance dilemma. A CM is defined as:

- (i) someone who is conditionally disposed to base her actions on a joint strategy or practice should the utility she expects were everyone so to base his action be no less than what she would expect were everyone to employ individual strategies ... (ii) someone who *actually* acts on this conditional disposition should her expected utility be greater than what she would expect were everyone to employ individual strategies.⁵⁰

There are three points to which Gauthier directs our attention. The first is that a constrained maximiser is one who possesses a *disposition* to behave in a certain way thus a constrained maximiser's conduct can potentially extend beyond immediate preference calculations. A constrained maximiser might find that 'in real interaction it is reasonable to accept co-operative arrangements that fall short of the ideal of full rationality' because this is the best that can be made of the situation – a CM therefore 'does not require that all acceptable joint strategies be ideal'⁵¹. As constrained maximisers operate in the real world, with imperfect rationality and knowledge, they accept less than fully rational outcomes so as to 'achieve mutually advantageous and reasonably fair outcomes'.⁵²

The second point is that a constrained maximiser's disposition is 'conditional on her expectation that she will benefit in comparison with the utility she could expect were no-one to co-operate'.⁵³ The constrained maximiser, therefore, calculates their expected utility based upon the 'likelihood that others involved in the prospective practice or interaction will act co-operatively compared to the outcome where no one co-operates – *she does not compare the outcome to one in which she defects*'.⁵⁴ Consequently, constrained maximisers do not leave themselves open to exploitation by a straightforward maximiser (SM) – a 'person who seeks to maximise his utility given the strategies of those with whom he interacts' – because she will adopt

⁵⁰ Gauthier, *Morals by Agreement* (n5) 167 (my emphasis).

⁵¹ Ibid 168.

⁵² Ibid 168.

⁵³ Ibid 169.

⁵⁴ Ibid 169.

straightforward maximisation if interacting with agents who are SM.⁵⁵ CM ‘makes reasonably certain that she is among like-disposed persons’ before she commits to herself to that disposition.⁵⁶ However, Gauthier does not consider situations in which CM agents may be ‘dealing with ... other sorts of agents’ beyond SM agents.⁵⁷

The final point, which arises from the disposition to constrain one’s direct pursuit of maximisation, is that ‘a constrained maximiser may find herself required to act in such a way that she would have been better off had she not entered into co-operation’.⁵⁸ This presents the greatest obstacle to the CM disposition, for CM requires that agents act in such a way that they *reduce* their payoffs given the range of actions open to them in a single interaction but this is only rational on the assumption that they *increase* their overall payoffs through the creation of opportunities unavailable to straightforward maximisers. Returning to the buffet example, if Edward, Patrick and George know that they will face the same interaction each working day then CM may be rational because it would increase the overall utility for each of them. Alternatively, the agent’s behaviour in a one-shot interaction would have to perform a signalling function to other agents who are not involved in that particular interaction for it to create additional opportunities for interaction – without this signalling function it would be irrational to comply. Conversely, in Bercovicz-type situations payoff maximisation can be achieved by exploitative non-compliance with the co-operative scheme *even* if this ends interaction. Gauthier’s dispositional approach attempts to show that even in a Bercovicz-type situation it is still rational *not* to maximise one’s payoffs by breaking the agreement.

Adopting Bercovicz-type behaviour is rational unless CM increases one’s opportunities to co-operate, thus increasing one’s overall payoffs and making it a rational disposition. The concept of a disposition ‘claims, any action that “expresses” this disposition *also* counts as rational’ – that is, rationality must be transitive.⁵⁹ The success of Gauthier’s project rests on the claim that it increases overall payoffs by increasing the opportunities for interaction and that, when accepting a lower payoff in an interaction, the lower pay off is rational when taken under the auspices of a rational disposition. If dispositions fail to do this then Gauthier fails to show that agents

⁵⁵ Ibid 167.

⁵⁶ Ibid 169.

⁵⁷ Danielson (n39) 295.

⁵⁸ Gauthier, *Morals by Agreement* (n5) 169.

⁵⁹ Copp (n33) 200.

should be constrained maximisers who comply with all co-operative agreements thus self-regulation by participating agents will be insufficient to ensure compliance and we will be forced to look elsewhere for a solution.

The Distinct Events Challenge

Holly Smith questions whether CM can close the compliance gap and suggests that the dispositional theory is reliant on a *perseverance principle* which entails that ‘[i]f it is rational for an agent to form the intention to do A, then it is rational for the agent to actually do A’.⁶⁰ Yet the gap between agreement and compliance remains because ‘intentions and the intended acts are two distinct events ... Hence the appropriateness of an intention appears to imply nothing conclusive about the appropriateness of the intended act’.⁶¹ The distinct events challenge denies there can be a ‘transitivity of rationality’ between the agreement to constraint and compliance, therefore, an action does not become rational because it is an act selected by a rational disposition.⁶² But this is exactly the claim of dispositions – an act becomes rational because a rational disposition selects that action. But even if it is rational to form the *intention* to act at t^1 and subsequently rational to *perform* the act at t^2 ‘we cannot use this fact to support the claim that the existence of her prior intention to do A *makes it* rational to do A, since it would be rational for her to do A whether or not she had ever formed that prior intention’.⁶³

The following example of a telepathic terrorist demonstrates the problem with Gauthier’s theory of dispositions:

[A] government official [is] negotiating with a terrorist. The terrorist, an infallible mind reader, threatens to blow up a plane load of innocent people unless the official forms the intention of releasing the terrorist’s imprisoned comrades. When, and only when, the official forms this intention (a psychological event the terrorist will detect telepathically), the bomb will be disarmed. Clearly, under these circumstances it would be rational for the official to form the intention of releasing the terrorist’s comrades. Once he has formed the required intention and the terrorist has

⁶⁰ Smith (n38) 244.

⁶¹ Ibid 245.

⁶² Sayre-McCord (n48) 183.

⁶³ Smith (n38) 245.

permanently disarmed his bomb, it seems equally clear that it would be rational for the official to change his mind and *not* release the comrades.⁶⁴

Under the theory of dispositions such a discrepancy could not occur. The official would still be (indeed must be) rational to release the terrorist's comrades. In order for dispositions to work this way it requires 'an extremely strong perseverance principle ... that prior intentions create *conclusive* reason[s] to carry ... [actions] out.'⁶⁵ Consider another stronger example against the strong perseverance principle:

[A] telepathic burglar threatens to steal all my household valuables. I know that if I form the intention of blowing up the house with the burglar and myself inside, it is nearly certain that he will be deterred. I form this intention, but unfortunately he is not deterred. According to the strong perseverance principle, it is now rational for me to blow up the house and kill myself, merely because I previously formed the intention of doing so under these circumstances.⁶⁶

We must therefore conclude that 'the rationality of the perseverance principle is false' – accordingly, 'Gauthier has not shown that it is rational to carry out constrained maximisation, even in those cases where it is rational to adopt it'.⁶⁷ The conceptual problem of perseverance arises because '[t]here are serious difficulties lurking behind this assumption of the transitivity of rationality. Most significantly, the transitivity cannot plausibly hold in all cases'.⁶⁸ This is a significant problem for the theory of dispositions because it means that it 'has not shown that it is rational to abide with moral rules once accepted, even if it is rational to accept them'.⁶⁹ This suggests that compliance needs a separate, distinct, rational justification in addition to the rationality of making an agreement – in other words, *compliance with constraint must itself be rational independently of any rational agreement*. One possibility is if agents prefer that they comply after making an agreement – in this case the solution would

⁶⁴ Ibid 246.

⁶⁵ Ibid 246.

⁶⁶ Ibid 247.

⁶⁷ Ibid 248.

⁶⁸ Sayre-McCord (48) 184, fn.6.

⁶⁹ Smith (n38) 248.

be psychological not rational because the agents changes their attitude towards compliance after agreeing to constraint.

Dispositions as a Psychological Mechanism

The only way in which CM could *always* be rational – and where compliance will always provides the highest payoff – is if ‘preferences will have shifted in tandem’ with the adoption of the CM disposition.⁷⁰ This avoids the problem of transitivity because the preferred outcome *changes* thus increasing the payoff from compliance while on the transitivity model the payoff is unaltered. But this conception of dispositions treats the gap between the rationality of agreement and compliance as a *psychological gap* because the CM reasoning does not include defection. If dispositions are a *psychological* solution this may indeed work – for exactly the same reason that psychological conditioning can ensure certain behaviours *but* this would not solve the problem of *rational compliance*. For ‘if the gap is not closed by the logic of [a] nonquestion begging conception of practical reason, it may be enough if it is closed, for the most part, by the psychology of [constrained maximisers]’.⁷¹ If this psychological solution is chosen – and preferences change upon the adoption of CM – then ‘compliance necessarily becomes the utility-maximising act’.⁷² This would solve both the distinct events challenge and the Foole’s claims regarding maximisation because at the point of compliance one’s preferences will have shifted to *include* compliance itself.

Yet, Gauthier rejects the possibility of psychological dispositions modifying one’s preferences because it would lead to the creation of a system in which ‘the preferences and capacities created by the social institutions and practices must be identical with the preferences and capacities best satisfied by those institutions and practices’.⁷³ Agents could justifiably be conditioned to have a particular set of preferences by the very institutions which provide satisfaction of those preferences. Furthermore, it would allow politically and socially dominant agents to condition others to have the same preferences – one might think of the conditioning that occurs in cults or far-right nationalist groups. In the case of constraint, compliance would condition agents to comply even when this would not maximise their utility. Aldous

⁷⁰ Copp (n33) 206.

⁷¹ Ibid 207.

⁷² Smith (n38) 248, fn.26.

⁷³ Gauthier, *Morals by Agreement* (n5) 323.

Huxley's *Brave New World* is 'the exemplar of a society designed to shape individuals so that they fit its institutions and practices'.⁷⁴ Such a system is '*artificially just*' because it ensures justice by conditioning each member of society to the social institutions, rather than adapting society to suit its members.⁷⁵

Consequently, dispositions cannot be interpreted as a psychological mechanism because this allows indoctrination and conditioning of one's preferences so that a brave new world society becomes the optimal outcome for interaction. Even if we think that no society could achieve such hegemony over the preferences of its subjects, which the history of totalitarian regimes suggests is impossible, we must still exclude this interpretation because it would encourage conflict between different ideological groups who would be justified in conditioning other agents on the grounds that this would ensure agents altered preferences would be maximised. Having rejected the possibility of preferences changing to include compliance, we can only rely on the idea that compliance in any particular instance of interaction is rational – even if sub-optimal – because of the strategic effect on subsequent ongoing interaction. Thus it is the strategies that agents adopt, not their dispositions, that we must be concerned with because disposition fail to ensure compliance and (because they are relatively fixed behavioural patterns) they leave agents open to exploitation. Agents using strategies can react to other agents with greater responsiveness and dynamism.

On a strategy account, agents should adopt CM because 'persons disposed to co-operation only act co-operatively with those whom they suppose to be similarly disposed', since 'a straightforward maximiser does not have the opportunities to benefit which present themselves to the constrained maximiser' SMs are not acting rationally.⁷⁶ If we know that we have to interact with the same agents in repeated co-ordination problems, those who comply (and who only comply with others who also comply) would benefit from the utility they accrue from *all* those interaction taken *collectively*. In these cases CM might be rational but it requires treating the collective payoff to compensate for the individual losses in separable rounds of interaction.

Without the transitivity of rationality between a disposition and compliance or the psychological interpretation of dispositions, Gauthier's solution fails because it

⁷⁴ Ibid 323.

⁷⁵ Ibid 323.

⁷⁶ Ibid 172.

requires that an agreement to constraint makes it rational to act in way that would otherwise be irrational when the only justified constraints are rational ones. A gap thus exists between forming an intention to act in accordance with constraint and the performance of that act at a later time – performance it contingent upon the compliance leading to a greater payoff in repeated interaction than non-compliance. But into this gap steps the Foole, who will renege on his prior intentions whenever it is no longer rational (in the maximising sense) to comply with constraint.

The Foole

The Foole is a straightforward maximiser who will break and trample upon any and all agreements if, by doing so, he can maximise his payoffs. The Foole makes his appearance in Hobbe's *Leviathan* arguing the following:

[T]here could be no reason, why every man might not do what he thought conduced thereunto [his preference maximisation] ... therefore also to make, or not make; keep, or not keep covenants.⁷⁷

The Foole 'insists that it is rational to co-operate only if the utility one expects from acting on the co-operative joint strategy is at least equal to ... one's best individual strategy'⁷⁸. If one could achieve a greater payoff through agreeing to and then failing to comply with an agreement then, the Foole argues, one should do so.

However, this 'defeats the end of co-operation, which is in effect to substitute a joint strategy for individual strategies in situations in which this substitution is to everyone's benefit' because it makes it no longer rational to agents to comply, eventually leading to the abandonment of the co-operative joint strategy and leading to a return to the state of nature.⁷⁹ After all, '[t]here is no sense in pursuing agreement on a cooperative joint strategy, if rationality precludes compliance'.⁸⁰ Note there is an asymmetry here – if compliance is known to fail this provides a reason *not to* agree to constraint but the making of an agreement *does not give* one a reason to comply.

⁷⁷ Hobbes (n14) 96.

⁷⁸ Gauthier *Morals by Agreement* (n5) 166.

⁷⁹ Ibid 166-167. Here, I assume that the Foole's strategy is known to the other contractors, I will return to this issue when discussing deception.

⁸⁰ Jody Kraus and Jules Coleman 'Morality and the Theory of Rational Choice' in Peter Vallentyne (ed) *Contractarianism and Rational Choice* (Cambridge University Press 1991) 258.

The problem of the Foole does not arise in Rawls's theory of justice because he assumes that agents in the original position (Rawls's initial bargaining situation) 'are not presumed to be egoistic or selfish'⁸¹, the 'rational individual does not suffer from envy'⁸², 'the parties do not seek to confer benefits or to impose injuries on one another' and 'their capacity for a sense of justice insures that the principles chosen will be respected'.⁸³ Given such assumptions there is no possibility of the Foole making an appearance but this is a weaknesses in Rawls' theory of justice, not a strength. The original position is so characterised that the Foole *cannot* make an appearance and this prevents Rawls from solving the compliance dilemma.

CM is presented as the solution to the Foole because a CM adopts the strategy *only comply with others who also comply* and so 'refuses to open herself to exploitation by acting cooperatively with an amoral straightforward maximiser'.⁸⁴ Because CMs co-operate there are some cases in which a CM does better than a SM and in these cases an agent should comply with constraint. In the famous Prisoner's Dilemma⁸⁵ a straightforward maximiser does worse then a constrained maximiser. Two people have been arrested. They can be charged with a lesser offence but their guilt in a more serious offence can only be proven by confession. If one confesses but the other does not, then the one who confesses will receive a much shorter sentence than the other. If both confess, they receive moderately longer sentences and if neither confesses, they receive a shorter sentence, albeit not as short as the single confession. This leads to the following payoffs:⁸⁶

	He confesses	He does not confess
I confess	5 years each	1 year for me, 10 for him
I do not confess	10 years for me, 1 for him	2 years each

Each individual (acting on SM) can do best if they confess and the other does not. Adopting an SM strategy each reasons as follows:

⁸¹ John Rawls, *A Theory of Justice* (Harvard University Press 1999) 111.

⁸² Ibid 124.

⁸³ Ibid 125.

⁸⁴ Danielson (n39) 294.

⁸⁵ Gauthier attributes the origin of this example to A. W. Tucker, see Robert Luce and Howard Raiffa (eds), *Games and Decisions* (John Wiley and Sons 1957) 94; see also Gauthier, *Morals by Agreement* (n5) 79-80.

⁸⁶ Table from Gauthier, *Morals by Agreement* (n5) 80.

If he confesses, then I had better confess – otherwise I’m in for ten years. If he doesn’t confess, then if I confess I’m out in a year. So whatever he does, I should confess. Indeed (each being a student of strategic rationality and the theory of games) in this situation there is but one outcome in equilibrium – the product of mutual confession.⁸⁷

The result is that each gets five years in prison. However, we can see that they could each have had less than five years – two years – if neither confessed. Mutual confession is the result of ‘mutually utility-maximising rather than mutually optimising strategies’ thus ‘supposedly rational utility-maximisers [SMs] do much worse for themselves than could supposedly irrational optimisers [CMs]’.⁸⁸ This is why a CM can do better than a SM – two CMs will remain silent thus receiving only two years each. Here, being a CM leads to the optimal outcome but the Foole will still propose defection because this generates an even greater payoff. If this is only a one-shot interaction it is unclear how to ensure compliance. In repeated interaction CM can punish SM by refusing to interact with them but this option is not available here. Thus, by refusing to co-operate with SM, the CM avoids the worst outcome of 10 years and so does the best she can in these circumstances.

If this refusal strategy holds together, then a CM strategy will ‘escape the compliance dilemma and show moral constraint to be individually rational’ because an SM strategy will *always* do worse against a CM strategy as CM will never co-operate with SM.⁸⁹ A constrained maximiser, by refusing to co-operate with the Foole, prevents him from having the opportunity to defect resulting in the Foole receiving only the payoffs of non-co-operation. The crucial part of this solution is that by excluding SM agents, we have good reasons to become constrained maximisers despite the fact that we may have to act in a way that is sub-optimal because this ensures our participation in continual interaction.

However, this is to rely on an external sanction to make compliance rational – if we do not comply then others will refrain from co-operating with us and we will suffer for it. It is the threat of *exclusion* that makes compliance rational not the

⁸⁷ Ibid 80.

⁸⁸ Ibid 80.

⁸⁹ Danielson (n39) 294.

disposition possessed by the agent. This would still not make it rational to comply in a single round of interaction nor in Bercovicz-type circumstances. Gauthier's notion of disposition fails to close the gap between agreement and compliance because either dispositions require transitivity of rationality (which the distinct events challenge fatally undermines) or dispositions must psychologically shift preferences to include compliance (which has been rejected because it permits indoctrination and conditioning) or dispositions only apply when the payoff from co-operation is greater than *non-co-operation and defection* due to the increased opportunities for interaction it provides. Only in the latter case is constrained maximisation a rational strategy but this crucially relies on agents being transparent, it does not take account of the high information costs associated with knowing whom to trust or the possibility of deceptive behaviour.

Translucency and Rational Deception

The basis of the rationality of adopting the CM strategy⁹⁰ is that one thereby increases one's opportunities for productive interaction and thus gains a greater payoff than an SM agent. This suggests that knowledge of the kind of agent one interacts with is crucial to determining whether CM is rational. Thus one form of the 'compliance dilemma is caused by a failure to discriminate. Unconditional co-operators fail to protect themselves from exploitation by straightforward maximisers' – protection from exploitation requires discrimination between agents, which entails conditional co-operation.⁹¹ As we saw, if an agent increases the opportunities for productive interaction it will be rational to comply even in sub-optimal cases:

Once the benefits of cooperation are taken into account, it seems that in choosing one's character, self-interest itself recommends abandoning self-interested reasoning in favour of moral reasoning.⁹²

The problem is that when 'facing equal opportunities, an enlightened egoist [the Foole of old] may expect to prosper'.⁹³ As David Hume puts it:

⁹⁰ We will leave aside, for the moment, the issue of multiple strategies and will assume that the strategy options of agents are only between SM and CM.

⁹¹ Danielson (n39) 297.

⁹² Sayre-McCord (n48) 187.

⁹³ Ibid 186.

[H]onesty is the best policy, may be a good general rule, but is liable to many exceptions; and he ... conducts himself with most wisdom, who observes the general rule and takes advantage of all the exceptions.⁹⁴

Given equal opportunities, rational constraint would indicate that people should be wise Humean enlightened egoists – an agent who exploits all others when they can do so without jeopardising their participation in interaction. Yet, such behaviour is barred by the ‘transparency assumption [which] would rule out, by fiat, the possibility of deception’.⁹⁵ Transparency assumes that agents know what disposition or strategy is possessed by all the agents with whom they are interacting. The more realistic assumption of translucency assumes that the likelihood of agents ‘to co-operate or not may be ascertained by others, not with certainty, but as more than mere guesswork’.⁹⁶ The issue then depends upon the opportunities afforded for CMs in a population of translucent agents. There is a further more serious problem, if all that is needed to defeat constraint strategies is that one has access to equal or greater opportunities then all one needs to do is *deceive* others into providing those opportunities. This can be done by being opaque rather than translucent. This is a significant problem because our reply to the Foole relies on one being ‘a (sufficiently) translucent member of a community of (sufficiently) translucent moral people’.⁹⁷ That is CM requires that other CMs can be identified so that exclusion of SM is possible.

An agent is ‘*opaque* if and only if others have merely an equal chance of correctly identifying her character ... [while an agent is] translucent if and only if others have better than equal chance of correctly identifying her character’.⁹⁸ As ‘translucency appears to be a realistic assumption concerning the extent to which real people are able to identify each other’s character’, the surety needed for the greater range of opportunities seems to be hanging by a thread.⁹⁹ This thread snaps entirely if ‘deceptive people will be careful to provide the requisite (though misleading)

⁹⁴ David Hume, *Enquires Concerning Human Understanding and Concerning the Principles of Morals* (first published 1751, Oxford University Press 1975) 282-3.

⁹⁵ Sayre-McCord (n48) 187.

⁹⁶ Gauthier, *Morals by Agreement* (n5) 174.

⁹⁷ Sayre-McCord (n48) 191.

⁹⁸ *Ibid* 188.

⁹⁹ *Ibid* 188.

evidence for those with whom they interact'.¹⁰⁰ This is because such 'people *seem* both translucent and trustworthy' with the consequence that 'their companions will quite reasonably, given their information, misjudge character with regularity'.¹⁰¹ As noted above, Gauthier's theory 'ignore[s] both the availability and the impact of deception' which can eliminate the opportunity-advantage of the CM strategy.¹⁰² Nevertheless, deception will not be utility-maximising if the costs of maintaining the deception strategy outweigh the gains made from its adoption. Essentially, this is the problem of high information costs that agents in a heterogeneous and anonymous population face when deciding whether co-operation and compliance with constraint is rational.

Thus, 'being opaque *may* severely limit one's prospects for cooperation'¹⁰³ if others detect that one is being opaque (rather than detecting deception directly) for they may be reluctant to co-operate (by erring on the side of caution), in which case the cost/benefit would firmly fall on the side of non-deception. Detection of opacity means 'your apparently innocent behaviour is taken by me to be *prima facie* evidence of the extreme duplicity of which you are capable', this in turn leads to 'the end of any possible civilized intercourse'.¹⁰⁴ To avoid this outcome, deceptive individuals must act so as to appear to be translucent, a concept that Sayre-McCord calls 'transopacity'.¹⁰⁵ If this is a viable strategy then 'transopaque enlightened egoists ... will enjoy many of the benefits and none of the burdens of being constrained by morality (assuming it is cheaper to maintain transopacity than it is to forego exploitation strategies)'.¹⁰⁶ The Foole thus remains a present threat to rational constraint but now armed with a deceptive strategy to remain undetected in order to exploit those who comply with constraint.

The success of this strategy, however, depends on the 'assumption that ... the costs of transopacity' are not raised by others.¹⁰⁷ Sayre-McCord finds this possibility unlikely, stating that punishing transopacity 'assumes (implausibly) that we live in a community that will raise significantly the risks of deception' by increasing the costs

¹⁰⁰ Ibid 192.

¹⁰¹ Ibid 192.

¹⁰² Ibid 191.

¹⁰³ Ibid 192.

¹⁰⁴ Narveson (n44) 172.

¹⁰⁵ Sayre-McCord (n48) 192.

¹⁰⁶ Ibid 193.

¹⁰⁷ Ibid 193.

of unsuccessful deception sufficiently to make non-deception rational.¹⁰⁸ This stance is hard to take seriously given that societies deploy and impose punitive sanctions for fraudulent conduct as well as punishing duplicity through mechanisms such as reputations, ostracism and refusal to interact with those whom we suspect to be untrustworthy. The investment governments make in pursuing fraudsters are specifically aimed at increasing the costs of exploitative actions. Of course, this does not mean that transparency cannot be successful even under these high risk conditions but it does seem that the threat of Sayre-McCord's deceptive maximiser (DM) strategy is somewhat overstated given that sanctions are routinely put in place which aim to make the costs of deception outweigh the benefits of exploitation. This would make it *irrational* to adopt DM – it is also possible that the use of sanctions could be extended to other exploiter strategies such as SM.

This is the first appearance of second-order constraints which modify the payoffs of particular strategies to make the rational or irrational. It is the presence of second-order constraint that leads to optimal outcomes. For example, a DM strategy cannot exist in a population made up only of other DMs or DMs and SMs (as they would all seek to exploit each other and prevent any successful co-operation or any successful exploitation). DMs, therefore, need to remain in a population predominantly made up of exploitable agents in order to be able to maximise their utility. But once within a population of CMs only second-order constraints can maintain the system of constraint and co-operation because by the time DMs become predominant in a population (thus provoking the withdrawal of CMs) co-operation will be collapsing. Second-order constraint can prevent or at least reduce the success of DMs without waiting until the point of no return. The role of second-order constraints also becomes more important as the number of strategies increases. When there are only two available strategies it is easier to identify what strategy other agents have adopted but the more strategies available the harder it is to identify agents. So how many strategies are there and how can we determine which one to adopt?

A Profusion of Strategies

As previously mentioned, the superiority of CM over SM is dependent upon the assumptions of transparency and the availability of only two strategies:

¹⁰⁸ Ibid 193.

The ... problem is that Gauthier mistakenly assumes both that (a) my only options are CM and SM, and (b) that my partner's only options are CM and SM.¹⁰⁹

But we have already seen that more than two strategies are available and, with but a moments thought, we can envisage many others. Consequently, 'there is no reason to suppose that our options are limited in this fashion' to two strategies.¹¹⁰ In Gauthier's defence, his definitions of straightforward and constrained maximisation are broadly defined such that they *are* the only two options:

*[C]onstrained maximisation [is] maximising our own utility subject to the constraint that we keep rational agreements ... [while] straightforward maximisers [maximise their] ... utility, even when it involves breaking a rational agreement.*¹¹¹

On these definitions CM and SM are the only two strategies because one either accepts constraint (thereby complying with rational agreements) or one does not. Using these definitions the acceptance of *any* conditional constraint counts as CM. However, the specific details of CM mean that there are a variety of alternative strategies that can fall under the umbrella of constrained maximisation – 'there are more dispositions to consider than just straightforward maximisation and constrained maximisation'.¹¹² For one strategy to be preferred to all other strategies it must be the uniquely rational one.

Let us consider a few more variations of strategies. Consider reciprocal co-operation as a possible solution and improvement over CM – '*Reciprocal Cooperation (RC): Cooperate when and only when cooperation is necessary and sufficient for the other's cooperation*'.¹¹³ Reciprocal co-operators 'only constrain themselves when cooperation is a *necessary* condition for the other's cooperation.'¹¹⁴

¹⁰⁹ Smith (n38) 238.

¹¹⁰ Ibid 238.

¹¹¹ Vallentyne (n23) 10.

¹¹² Copp (n33) 221.

¹¹³ Danielson (n39) 299.

¹¹⁴ Ibid 299.

Therefore, ‘RC cooperates with CC [Conditional Co-operators¹¹⁵] and RC itself and defects with SM and UC [Unconditional Co-operation]’.¹¹⁶ ‘RC is the best response to the other player’s best response to it [because they either comply or are excluded]... Thus, RC is both morally and rationally [superior to CC].¹¹⁷ However, from this definition of RC we can see that RC exploits UC through defection which maximises their utility.

Unconditional co-operators innocently trust all other agents to comply with their agreements but ‘[e]xploitation of the innocent appears to be morally indefensible’, yet it is this feature that makes RC *more* rational than CC because ‘one never does worse and sometimes does better using RC’.¹¹⁸ One, thus, maximises one’s payoffs to a greater degree than CC because defection in interaction with UC gives a higher return than co-operation. But Peter Danielson thinks that ‘RC, by making UC irrational, prevents exploitation of these innocents’ because RC forces innocents (UC) to adopt – or to learn – a more discriminating strategy for interaction.¹¹⁹ RC effectively makes UC too costly to pursue, while CC allows UC to flourish when a sufficient proportion of the population uses UC, this allows them to be exploited by SM. We can see that once we delve deeper into the options available to agents numerous strategies present themselves in numerous combinations.

But this has uncomfortable implications for it implies that the exploitation of universal co-operators to be rational – ‘[m]ust rational morality choose between two evils: CC, which abandons innocents, and RC, who preys upon them’?¹²⁰ In order to solve this *without resorting to second-order constraints* we need more developed strategies that would redress this problem of abandonment or exploitation. Enter the ‘*Unconditional Cooperator Protector (UCP): Cooperate with and only with those who (a) cooperate with Unconditional Cooperators and (b) Cooperate with Unconditional Cooperator Protectors*’.¹²¹ This strategy is ‘a great improvement, both morally and rationally ... [because] UCP protects ... unconditional cooperators ... [it

¹¹⁵ A Conditional Co-operator is Danielson’s equivalent to Gauthier’s CM; ‘Gauthier’s principle of constrained maximisation (CM) embodies such a conditional disposition ... Simplifying somewhat, I identify this component of constrained maximisation as the principle (or strategy) of conditional cooperation’ Ibid 297.

¹¹⁶ Ibid 299.

¹¹⁷ Ibid 300.

¹¹⁸ Ibid 300.

¹¹⁹ Ibid 319.

¹²⁰ Ibid 320.

¹²¹ Ibid 320.

is] morally superior'¹²² to both CC and RC because '[r]ational moral agents should try to make their world safe for morality *and innocence*, and this is not irrational'.¹²³ If it is rational to protect innocent agents, then UCP will be superior to RC. The difficulty with this response is that it seems to assume co-operators should be concerned about protecting the innocent but this would only be rational if by protecting them UCP increases the overall co-operative return beyond that achievable by RC. Alternatively, if an agent's preferences include the wellbeing of others a UCP-like strategy might be selected because this will protect those who are the object of an agent's preferences but we clearly cannot rely on such sympathy or concern to solve the compliance dilemma.

This is not the end, there are additional strategies available by dividing CM into *broad* compliance and *narrow* compliance. This division is based upon the following distinction: a broad complier is 'a person who is disposed to co-operation in ways that, followed by all, merely yield her some benefit in relation to universal non-co-operation'.¹²⁴ A narrow complier on the other hand is 'a person who is disposed to co-operate in ways that, followed by all, yield nearly optimal and fair outcomes'.¹²⁵ We can further divide narrow compliance into two more varieties. First, is the 'disposition *to agree and comply with fair bargains only*' while the second is the strategy '*to comply with whatever fair bargains they strike, though they have no particular disposition with respect to the striking of bargains, fair or otherwise*'.¹²⁶ Thus narrow compliers who agree to and comply (Nc/AC) with fair bargains will only interact with those who offer fair bargains, while narrow compliers who only comply with fair agreements (Nc/C) make any agreement that offers them an advantage but only comply with those bargains that are fair.

For those who adopt Nc/C, a decision to 'strike a particular bargain will be made on straightforward utility-maximising grounds' while Nc/AC's 'just do not make unfair bargains'.¹²⁷ The second version of narrow compliance (Nc/C) is the most rational because a 'more complex disposition that disposes agents to both comply and accept bargains is unmotivated for rational agents'.¹²⁸ In the context of

¹²² Ibid 320.

¹²³ Ibid 321.

¹²⁴ Gauthier, *Morals by Agreement* (n5) 178.

¹²⁵ Ibid 178.

¹²⁶ Kraus and Coleman (n80) 272.

¹²⁷ Ibid 273.

¹²⁸ Ibid 274.

successful co-operation, Nc/C's are able to protect themselves from exploitation because they decide on entering into bargains on straightforward maximisation grounds. One of the factors an Nc/C will consider is the likelihood of compliance by others, thus 'if the compliance of others is likely and the bargain struck a fair one, the narrowly compliant person is compelled to cooperate'.¹²⁹ Consequently, '[n]arrow compliance ... ensures the exclusion of individuals with unfair entitlements from bargains with narrow compliers' because narrow compliers are not bound to comply with unfair agreements.¹³⁰

These are just some of the possible strategies that are available but it is clear that for agents to accept and comply with constraint they must make their conduct conditional on the conduct of others and be prepared to expend resources to exclude or punish those who defect from co-operative constraints. The key to preventing exploitative predator-strategies is to make them too costly either by the behaviour of agents or by introducing additional costs. Evolutionary game theory will help explain how different strategies can prevent straightforward maximising behaviour while the notion of secondary constraints can adjust the payoffs of actions.

The Evolutionary Game Theory Test

To identify the most effective strategy for maximising preferences Danielson proposes an evolutionary strategic test, which analyses populations of different behaviours. Each agent 'must view strategic choice both as a response to the choices of his fellows and as being responded to by those choices'.¹³¹ This provides a 'fundamental justification ... [because] these crucial variables ... [i.e. the strategies of others] ... [cannot] be exogenously determined'.¹³² We cannot assume that the only strategies present are co-operative strategies. (This is distinct from assuming agents have pre-determined preferences, which concerns their goals and motivations not the methods they use to achieve them.) An agent 'can hardly fail to have some awareness of the fact that this environment is made up in part of other agents similar to himself'.¹³³

¹²⁹ Ibid 273.

¹³⁰ Ibid 277.

¹³¹ Gauthier, *Morals by Agreement* (n5) 60.

¹³² Danielson (n39) 302.

¹³³ Jon Elster, *Ulysses and the Sirens* (Cambridge University Press 1984) 117.

An evolutionary test measures strategies against each other by the frequency of their survival and resistance to invasion by other strategies. But ‘once agents are allowed to react to other’s ... complexity explodes and there are infinitely many agents [and infinitely many strategies] to consider’.¹³⁴ However, as agents are unable to grasp every possible strategy available to them this may not be a significant problem because any particular agent would, realistically, only be able to choose from a smaller set of strategies given time limitations in practical reasoning. Nevertheless, given this modelling system, there is in principle no reason why one could not model thousands of strategies – albeit perhaps through a sophisticated software program and significant computational resources. Such a test is agent focused, ‘[t]his test is evolutionary rather than ecological because it starts with the agent to be tested and assumes that new agents come into existence’.¹³⁵ An evolutionary test allows one to model the success – and thus the rationality – of at least some possible strategies within a dynamic population. Consider the following scenario:

In a population of retaliators, no other strategy would invade, since there is no strategy that does better than retaliator itself. However, dove [Unconditional Co-operator] does equally well in a population of retaliators. This means that, other things being equal, the numbers of doves could slowly drift upwards. Now if the numbers of doves drifted up to any significant extent, prober-retaliators [who test the reactions of others] (and, incidentally, hawks and bullies [straightforward maximisers]) would start to have an advantage, since they do better against doves than retaliators do [because SM exploits UC].¹³⁶

This example shows that a strategy can be tested in a strategic context against other strategies – retaliator is not a stable strategy because it allows the number of unconditional co-operators to increase, which in turn increases the payoffs of straightforward maximiser predator-strategies.

Thus, a stable equilibrium outcome may consist of a mix of strategies some of which might be predator-strategies. In a stable equilibrium there would be no internal

¹³⁴ Danielson (n39) 301.

¹³⁵ Ibid 304.

¹³⁶ Richard Dawkins, *The Selfish Gene* (Oxford University Press 2006) 74-5.

shift in strategy mix but this mix of behaviours may not be the most optimal because this would permit some predator-strategies to operate. Reputation and exclusion from future interaction may be insufficient if predator-strategies act deceptively. To increase the efficiency of co-operation agents should introduce second-order constraints that increase the cost of some strategies while increasing the benefits of others. Moreover, it may not be rational for an individual agent to retaliate because the costs of enforcement may be greater than benefits from having a reputation as a retaliator. In these cases publicly funded and enforced second-order constraints can fill the gap. The resources expended in criminal law are greater than any agent or even groups of agents could afford but through public funding by *all* (or nearly all) agents.

Given that agents must react to multiple strategies Danielson proposes the meta-strategy – a meta-strategy is ‘a strategy defined over others’ strategies’¹³⁷ – of an ‘indirect maximising (IM) agent, who can in *each* interaction choose that principle that will best satisfy here preferences’.¹³⁸ An IM strategy would select whatever strategy would be most beneficial in a given population because IM agents are ‘capable of selectively copying other’s public-constraining principles can implement reciprocal cooperation’ or adopt whatever strategy would work best.¹³⁹ Essentially, IM’s would perform an evolutionary test in order to see what strategy would be successful amongst the other known agents – agents ought to select that strategy as their interaction strategy.

The evolutionary test is, however, limited in one significant way. Whilst it can test strategies against each other, it can only direct us to adopt the strategy that maximises the payoff an agent receives. It cannot though tell us to adopt UCP as a rationally superior strategy to RC, because RC could invade a population of UCPs who allowed too many UCs to be present. Thus, a determination that innocence should be protected cannot be made based upon the evolutionary test. Furthermore, it would tell us to be deceptive by appearing to be a co-operator when in fact we are a straightforward maximiser if this would be the most effective strategy to adopt.

Under an evolutionary test, if a deceptive maximiser is the most successful strategy it should be adopted but such a strategy would be the very antithesis of an agreement amongst agents. It seems likely, however, that the DM strategy would only

¹³⁷ Danielson (n39) 307.

¹³⁸ Ibid 314.

¹³⁹ Ibid 314.

work in a population with other non-deceptive strategies, as it could not co-operate with itself. But the point remains that there may be a predatory strategy which would be more successful and rational than any other which an evolutionary test would indicate we should adopt. An evolutionary test is thus a useful analytical tool for assessing the viability of various strategies in populations without second-order constraints but the presence of even a few predator-strategies leads to sub-optimal outcomes because of the costs agents would take to protect themselves. The evolutionary game theoretic approach is therefore limited because it relies on self-enforcement by the population of agents. Yet, an even better outcome can be reached if predator-strategies can be eliminated entirely – or almost entirely – through second-order constraints.

Second-Order Constraint and a Rational Constraint Theory of Law

Agents will adopt the strategy which best fulfils their preferences but such choices will affect the strategies adopted by others thus determining what strategies are successful. Consequently, constraints will shift and change depending on the mix of strategies within the population. However, self-regulated constraint created by the strategies of others permits some exploitative predator-strategies to exist preventing an optimal system of constraints – this is why the use of second-order constraints becomes necessary. The notion of second-order constraint applied to legal theory results in a conception of law as a means of modifying the payoffs and costs of certain strategies to produce the most optimal outcome. In other words, legal rules are a behavioural tools whose function is to enhance co-operation by attaching penalties to certain actions and/or attaching rewards to others. By ensuring that agents comply with the dictates of constraint co-operation can be maintained and modified to produce the most optimal outcome for a given population of agents. We can develop second-order constraints which stabilise the viability of co-operative strategies and reduce even further the success of predator-strategies. For example, a population of agents might try and eliminate certain predatory strategies through the creation of devices such as a judicial regime, a codified constitutional structure, a Nozickian protective association or some other system of second-order constraint.

Where the costs to the agent themselves exceed the gains of enforcement – including the benefits of later interaction – second-order constraints step in to make sure the penalty is enforced or to create a penalty where none was present before. This

is desirable to prevent the proliferation of predator-strategies. If predator-strategies become prevalent a downward spiral ensues which will result in the breakdown of co-operation because agents are more likely to encounter defectors and will begin to act defensively against the greater frequency of predator-strategies. Second-order constraints provide a means, because they can apply universally to all interaction, of reducing the success of predator-strategies thus preventing them becoming prevalent and leading to the end of co-operation. This provides a novel way of analysing legal rules, not in terms of notions of natural law, principles, moral systems or other notions of objective law but in Legal Realist terms of social and psychological phenomena, the preferences of actual agents and the identification of optimal outcomes through rational choice. By linking legal rules to the preferences of agents, legal subjects become the beneficiaries of a judicial system designed *for them* rather than being passive colonised recipients of judicial power. The use of second-order constraints to preserve and enhance the benefits of co-operation is, therefore, justified by the benefits of co-operation and the reduction in predator-strategies allowing agents to pursue their goals more effectively. Additionally, second-order constraints can also mitigate the problems of weakness of will, hyperbolic discounting and the irrationality of the passions.¹⁴⁰

One difficulty that remains is how to distinguish between justified and unjustified second-order constraints. We might expect that a theory of rational constraint should be able to distinguish between the uses of force in a State from the use of force in criminal organisations such as the Cosa Nostra. For example, criminal fraternities impose a second-order constraint in the form of a threat on snitches by inflicting extreme violence upon those who break the behavioural constraint of silence.¹⁴¹ This is beneficial for those participating in the criminal enterprise but we would want to distinguish these constraints from legal constraints. But both sets of constraints benefit their participants and both use force as the mechanism for ensuring compliance – it is not therefore clear how this distinction could be made necessitating further work. One avenue for maintaining this distinction may be that claim that legal rules take into account all agents in a given population whereas the criminal fraternity only takes its own members into account. Yet such an approach would require that all

¹⁴⁰ See: George Ainslie, *Breakdown of Will* (Cambridge University Press 2001) and Elster, *Ulysses Unbound* (n11).

¹⁴¹ I am grateful for Prof. Søren Holm for pointing out this example of second-order constraint.

those who are involved in interaction are included which, in our globally integrated world, might require some planetary-level world state form of constraint because nation states are unable to account for all those who interact with the state organisation.

The existence of criminal offences (such as fraud and embezzlement) and societal sanctions (such as notions of reputation and honour) are exactly the sort of second-order constraints that agents would use to reduce the success of predator-strategies and increase the success of co-operative strategies. We could also consider third-order constraints that limit second-order constraints; for example constitutional systems (3rd order) which constrain legal regulations (2nd order) which, in turn, constrain individual actions (1st order). All these constraints are reducibly normative constraints because any agent who wants to fulfil their considered preferences *ought* to adopt a constraining strategy and they *ought* to comply with those constraints.

Conclusion

We can see that second-order constraint can resolve some of the problems associated with a pure system of self-regulated co-operation which relies on individual agents providing enforcement. In a system of self-regulation rational choice suggests that agents should act deceptively as long as the costs of deception are lower than the gains from exploiting other agents thus the costs to each agent of enforcement may exceed the benefits and so it would be irrational to punish an agent after exploitation has occurred *unless* future interaction is dependent on it – and even this may be insufficient if the cost is high enough. Amongst a small relatively homogeneous identifiable population such self-regulation is possible and, most likely, preferable because it is effective and easy to maintain. Once a large anonymous population of heterogeneous agents exists second-order constraint is needed to provide enforcement to sustain co-operation – primarily because of the high information costs that attend such populations. The use of the internet may help reverse this trend – for example, through the use of rating systems for businesses and products, and Facebook for individuals – but this will be limited to those who make no great effort to hide their strategic behaviour. Consequently, second-order constraints are still (and perhaps more) needed to make deceptive and exploitative predator-strategies irrational reducing even further their presence within the population of agents.

Self-regulation alone is insufficient because it lacks the strong perseverance principle of transitivity of rationality that Gauthier's seems to assume, it encourages agents to be deceptive and use predator-strategies when the costs are lower than the benefits. It was therefore necessary to consider how to improve upon Gauthier's original theory taking into account the critiques made by other scholars and the need to mitigate the effects of predator-strategies. A means of modifying the behaviour of agents in interaction was needed without modifying the preferences of agents themselves – because this would justify totalitarian tyranny and the forcible imposition of particular sets of preferences through the indoctrination of agents. For this reason second-order constraints were proposed as the solution to the problems plaguing self-regulation but this led to a revision of legal theory as the manifestation of rationally chosen rules to promote co-ordination. Implicitly, this treats law as a social and psychological phenomenon in the vein of Legal Realism which focuses on the practical effect of the decisions of judicial officials within the system.

Legal rules then are functional mechanisms for producing a better collective outcome for a population of agents – *better* being defined by the maximal satisfaction of an agent's preferences consonant with all other agents doing so as well. By modifying the cost and payoffs of certain actions laws can change an agent's behaviour without resorting to the constant use of force. More than this, however, by requiring legal rules (and constraints generally) to confer a benefit on those subject to legal regulation the system promotes the interests of all legal subjects and provides them a position from which to challenge laws that impose costs without benefit or law that imposes costs greater than the benefits received. Furthermore, this system takes agents as they actually are, not as a fictitious set of homogeneous reducible entities whose characteristics can be defined in a definitive list of metaphysical terms. The focus on the actual human condition and the actual preferences of agents implicitly makes law accountable to each agent without the need for moral agreement on moral equality, human rights or any other metaphysical characteristic.

Legal rules under this approach are functional devices for promoting co-operation by reducing the incentives that remain under a system of self-regulation for predation while benefiting all agents. By using the metric of benefit to justify constraint all agents can challenge the judicial system and can identify when legal rules become unjustified. I hope that through such a system the individual will be placed in a stronger position in relation the legal rules introduced by the institutions of

the State in which they live. Agents will be an active component in deciding the rules by which we all live reducing the homogenising dominance of particular conceptions whether of morality or society (for example, nationalistic conceptions of a socio-political unit). Under this revised theory of rational constraint agent should be able to achieve more of their goals, legal regulation can be limited to its necessary elements and judicial officials can be subject to greater scrutiny – taken together this will produce a system of constraint and co-operation that establishes the primacy of individuals over any one particular worldview.

Chapter Three

Is a Waiver of Human Rights a more Rational and Beneficial Theory of Rights?

Introduction

Human rights law ascribes rights to human individuals¹ which constrain the conduct of States towards those under their jurisdiction. Recently, however, constraints have been applied to inter-personal conduct between right-holders themselves thus limiting right-holders actions – a clear example is demonstrated by euthanasia cases under which an agent cannot grant permission for active euthanasia because, as we shall see, the right to life does not entail the power to waive the exercise of the right-based protections. Note that human rights create an asymmetrical negative right constraining the actions of others by allowing a right-holder to refuse certain treatment or otherwise limit what other agents can do to the right-holder – human rights do not create a positive right to request particular actions. Additionally, States can lawfully infringe most human rights² while third-party right-holders cannot. It is the existence of these asymmetries that leads to the constraint of a right-holder's actions because it denies them the power to choose how and whether their rights are exercised.

The mere presence of these asymmetries is not an argument against the current regime of human rights law – for they may be well justified – but from the rational constraint perspective adopted in this thesis we can ask: is the inability of right-holders to waive the exercise of their rights necessary for securing co-operation? For individual agents 'will not merely be the intended *beneficiaries* of these [human rights] obligations, but will carry the intended *burdens*' because the 'second-order obligations of states are discharged by imposing first-order obligations on others and

¹ I am not concerned with *to whom* human rights are ascribed – rather my concern is with *how* those rights operate. The Universal Declaration of Human Rights and the European Convention on Human Rights are, perhaps, two of the most well known human rights documents. Note that the legal position is that only born humans, members of the species *homo sapiens*, possess human rights. See: Paul Sieghart, *The International Law of Human Rights* (Clarendon Press 1983) 130 where he states that 'human rights can only attach to living human being' and the judgement in *Vo v France* (2005) 40 E.H.R.R. 12.

² Paul Craig and Gráinne de Burca, *EU Law* (5th edn, Oxford University Press 2011) 526.

enforcing them'.³ This chapter seeks to challenge the constraints imposed upon right-holders by the judicial human rights system.

But what if these constraints are justified? On the theory of rational constraint is justified only if it is necessary to secure co-operation and confer benefits on agents – if an alternative system is possible that can achieve this without constraint or with a lesser constraint then it is not justified because it imposes unnecessary costs upon agents. The challenge articulated here is that constraining the actions of right-holders – by denying them the power to change how and when their rights impose duties on others – is unjustified because this limitation is not necessary for securing the beneficial protections that the human rights system creates. To remove the application of these constraints to the right-holder we need to consider whether it makes sense for a right-holder to be able to forgo their rights and duties in some way.

Of particular interest are the constraints imposed under the European Convention of Human Rights and the right to found a family (Art.12), the right to private and family life (Art.8) and the right to life (Art.2). All the cases discussed in this chapter illustrate ways in which the actions of right-holders have been limited by these rights and is the unifying theme of the cases analysed here. If human rights permit an agent to waive the protections of those rights in a particular instance (for example, allowing a surrogate to waive their status as the legal mother of the offspring) – *without giving up the possession of the right* – by including voluntary waiver in our conception of inalienability, then the restrictions that human rights impose upon right-holders would be minimised. The justification for reducing the constraint imposed upon right-holders is as follows – agents will be better placed to fulfil their preferences when their actions are only limited to the minimal degree necessary for co-operation. It will, therefore, be argued that in relation to reproduction it will be more beneficial for agents if they can waive the exercise of their rights.

This analysis of the concept of inalienability will utilise Joel Feinberg's parsing of inalienability into a number of distinct forms and Wesley Hohfeld's work on the relations that constitute rights explaining how a revised sense of inalienability might work. Inalienability can thus be refined and understood as the irrevocable possession of a right itself but the exercise of which can be waived in relation to the specific actions, whilst retaining its limiting force in relation to States. Feinberg's

³ Onora O'Neill, 'The Dark Side of Human Rights' (2005) 81 *International Affairs* 427, 436.

work enables a distinction between retaining the *right* to X while waiving X (the good the right secures) and Hohfeld's analysis explains the mechanics of how a legal capacity based on this would work. For example, I have a right to found a family but I waive that right if I donate to a sperm bank (at least in the case of that instance of donation) because I forgo the right to decide *how* those sperm are used, the responsibility for any offspring produced and the right to perform the role of a father to the offspring. A similar system is proposed for human rights law. An explanation is required of two preliminary issues before we begin. First, the ambiguity between *legal* and *moral* human rights, and second, the reason that right-holders should only be minimally constrained in the context of human rights.

Moral and Legal Human Rights

Discussion of human rights is susceptible to significant ambiguity between legal and moral conceptions. Allen Buchanan scathingly criticises the perpetuation of this ambiguity.⁴ While 'international human rights law is the universally accessible *authoritative* version of the global moral lingua franca'⁵ many assume 'that there must be a corresponding, antecedently existing moral human right' grounding legal human rights.⁶ Thus, the 'genesis of human rights, on the standard view, is not in a prior voluntary act'.⁷ Similarly, human rights 'demand ... the social changes required to realise the underlying moral vision'.⁸ These claims do not seem possible if human rights are moral claims.

Buchanan calls this the Mirroring View and illustrates the fallacy of this assumption in relation to the international human rights legal system. The problem with the assumption underlying the Mirroring View is that some attempts 'to enforce a moral right would ... require dangerously high levels of coercive capacity on the part of state' and, consequently, 'not all moral rights are suitable candidates for legalisation'.⁹ Buchanan rejects the Mirroring View because there would either need to be a one-one mapping between moral and legal rights or, alternatively, a single

⁴ Note that Buchanan primarily discusses the *international* human rights legal system, whilst his arguments relating to the role of moral rights and moral justification in the human rights law is relevant, the international dimension (in the sense of State-to-State conduct) will not be analysed in this work.

⁵ Allen Buchanan, *The Heart of Human Rights* (Oxford University Press 2014) 7.

⁶ *Ibid* 14.

⁷ Donald Veer, 'Are Human Rights Alienable?' (1980) 37 *Philosophical Studies* 165, 166.

⁸ Jack Donnelly, *Universal Human Rights in Theory and Practice* (Cornell University Press 2003) 15.

⁹ Buchanan 56.

determinate set of moral principles that human rights law is to express.¹⁰ Furthermore, the existence of a plurality of moral systems – due to all cognitivist beliefs in objectively existing moral facts being false (see chapter one) – leaves open the possibility of multiple moral human rights systems. Human rights should be seen as a non-moral legal device justified by the theory of rational constraint which promotes co-operation by constraining actions and securing an agent’s control over their lives, traditionally against the State. The concept of rational constraint means we can develop a system of human rights law that promotes the fulfilment of preferences to a greater degree than the current regime without undermining either the protections such rights confer or the co-operative scheme as a whole.

On this understanding my focus is on the legal human rights systems, not human rights as a moral right. For jurisdictional reasons, I am specifically concerned with the human rights regime that has developed under the auspices of the European Convention on Human Rights (the Convention) and the European Court of Human Rights as incorporated into English law by the Human Rights Act 1998. The Human Rights Act came into force in 2000 but indirectly incorporates the Convention rights into English law requiring that all primary and secondary legislation is interpreted (so far as possible) to be compatible with the Convention and makes it unlawful for a public authority to act in an way that is incompatible with the Convention (see Human Rights Act 1998 ss6-8). The Convention, through the Human Rights Act, regulates the conduct of the State and its departments *but* it does not directly regulate the conduct of individuals. The constraints imposed arise from the judicial regime’s interpretation of how these rights operate and the concepts they use to justify their decisions.

A secondary issue concerns justifying the constraints agents must abide by. As I argued previously (in chapter two), only a theory of rational constraint can justify constraint because it does not rely on a moral premise instead using the interests and goals of agent themselves to justify constraint. Agents accept constraint only when it is rational to do so because it promotes the fulfilment of their preferences. For example, criminal law protects all agents from widespread random violence thus removing the need to personally invest resources in defence and protection. It would not be rational for any agent to be constrained by the criminal law if others were not

¹⁰ Ibid 14-21.

similarly constrained.¹¹ Moreover, different agents have disparate goals and ‘[n]umerous alternative imputations of value could be discovered or invented’¹², whilst the ‘notion of a perfect whole, the ultimate solution, in which all good things coexist ... [is] not merely unattainable ... but conceptually incoherent’.¹³ Coupled with the acceptance of moral error theory¹⁴ *moral justifications* for constraint are not possible but *rational justification* for constraint is possible because its practical means-end reasoning allows agents to best achieve their aims giving them a reason to accept constraint. Rational constraint treats human rights law as creating a system of rules to ensure co-operation amongst agents leading to greater preference fulfilment than in the absence of such an agreement. Essentially, we begin from the Hobbesian idea that human rights represent ‘the liberty each man hath, to use his own power, as he will himself’ with all constraint being justified by voluntarily forfeiting some of their power to secure co-operation.¹⁵

The claim that moral rights are prior to the existence of human rights law has been rejected for the reasons set out above. Yet, case law shows that there is a moral element within judicial reasoning. Lord Sumption, for instance, has stated in an end of life case:

English judges tend to avoid addressing the moral foundations of law. It is not their function to lay down principles of morality, and the attempt leads to large generalisations which are commonly thought to be unhelpful. In some cases, however, it is unavoidable. This is one of them.¹⁶

The role of moral claims in the judicial process, indeed in justifying constraint in all spheres of life, should be left to individuals and should not form any part of the general constraint under which social groups operate because moral claims are not objective truths about reality. In euthanasia cases the claim that it is immoral for medical professionals to bring about the death of their patients is merely a statement

¹¹ A recognised exception would those offices that empower the office-holder to use force (for example law enforcement agencies) while prohibiting the use of force against those officials.

¹² Loren Lomasky, *Persons, Rights and the Moral Community* (Oxford University Press 1990) 21.

¹³ Isaiah Berlin, *The Proper Study of Mankind: An Anthology of Essays* (Random House Publishing 1998) 11.

¹⁴ See: John Mackie, *Ethics: Inventing Right and Wrong* (Penguin Books 1990), Jonas Olson, *Moral Error Theory* (Oxford University Press 2014) and Chapter One.

¹⁵ Thomas Hobbes, *Leviathan* (first published in 1651, Oxford University Press 1996) 86.

¹⁶ *Regina (Nicklinson) and another v Ministry of Justice* [2014] UKSC 38 (Supreme Court) para.207.

of one particular normative viewpoint – it cannot justify constraint amongst a population with members who do not accept those moral claims. Consequently, the current human right legal regime prevents some from fulfilling their preferences when, as will be shown, this is unnecessary for co-operation.

Traditionally, human rights law exists as an international system (founded on the Universal Declaration of Human Rights) that limits a State's power over its own subjects. Human rights law is aimed at State governments – not individuals – by imposing 'obligations on many governments as to what they may or may not do to individuals over whom they are able to exercise State power'.¹⁷ Thus it would seem that restrictions on individual agents would not occur but:

Detailed control is needed to 'achieve progressively the full realisation' of very complex sets of potentially conflicting rights, which must be mutually adjusted ... Those who frame it [human rights law] have to seek to ensure that individuals and institutions conform to a very large number of constraints in all activities, so have to set and enforce very detailed requirements.¹⁸

For human right law to achieve its goals it must allow the State to exercise micro-level control over a much of an agent's life by prohibiting certain conduct. This may explain the expansion of human rights law to include an inter-personal function by prohibiting certain third-party interventions in addition to its traditional State-limitation function. As an illustration, in *Evans v the United Kingdom*¹⁹ a woman sought to have children with her former husband's sperm under Articles 8 and 12 of the European Convention of Human Rights and part of the justification given for not permitting Evans the use of her eggs fertilised with her ex-partner's sperm was his right not to become a father. This case was technically about gaining State permission for Evans to have the fertilised egg implanted in the absence of her partner's consent but the justification for not doing so was that her ex-partner would have become a father against his will, violating his human right to found a family, which entails (in

¹⁷ Sieghart (n1) 15.

¹⁸ O'Neill (n3) 436.

¹⁹ (2006) 43 E.H.R.R. 21.

this right at least) the opposite right *not* to found a family.²⁰ As Arden L.J. stated in the Court of Appeal, ‘it would amount to interference with the genetic father's right to decide not to become a parent’.²¹ It is interesting to note that one of the factors Arden L.J. considered relevant in reaching this conclusion was that the fact that ‘it will probably involve financial responsibility in law for the child’ – this is one of the key factors that would be easier to resolve if an agent waived their *parental rights and duties* (but this is to get ahead of ourselves).²²

Euthanasia cases have also restricted right-holder actions on grounds relating to third parties – the Supreme Court of the United Kingdom in *Nicklinson* approvingly quoted the Falconer Commission’s statement that ‘vulnerable people could be put at risk of abuse or indirect social pressure to end their lives, if such an option [assisted dying] was to become available’.²³ However, the ‘Falconer Report [also] indicated that in three jurisdictions where it was permissible to assist suicide, there was no evidence of vulnerable groups being subject to any pressure or coercion to seek an assisted death’ thus challenging the idea the coercion would be exerted.²⁴ The bare risk of coercion exists in any interaction and without stronger evidence that allowing particular actions will increase or inherently is coercive we should not impose that constraint on agents. The risk of coercion (if present) can be mitigated by the process itself while the mere risk of coercion is insufficient to justify constraining agents. Thus, orthodoxy holds human rights law as an other-regarding constraint binding States – right-holders do not seek to bind themselves – yet, human rights seem to have developed (albeit in a limited fashion) an additional function limiting what right-holders can do.

The Constraint of Human Rights Law

Why is it a problem if human rights law imposes constraint upon right-holders? Human rights are an integral part of discourse about interaction in legal, political and social arenas and one of the most important tools of interpersonal co-operation. While they are by no means an inevitable component of co-operative interaction²⁵, they have

²⁰ Ibid 413-5.

²¹ Ibid 415.

²² Ibid 415.

²³ Commission on Assisted Dying, ‘Final Report’ (2012) 287-8.

²⁴ *Nicklinson* (n16) SC Lord Neuberger of Abbotsbury para.54.

²⁵ Co-operation within social groups, obviously, occurred prior to the contemporary human rights legal system.

become a powerful means through which individuals assert some sphere of non-interference in their relationships with others. Human rights have become such powerful tools in asserting individual rights because they are considered to be universal, inherent and inalienable.²⁶

This makes them unique amongst legal rights which are generally only those rights that the system itself creates and confers. In contrast, human rights are explicitly designed to ‘impose obligations on many governments as to what they may or may not do to individuals over whom they are able to exercise State power’ – *even if the States legal system does not contain any human rights*.²⁷ The development of *international* human rights law justifies the imposition of duties upon State governments as a ‘wider reordering of the normative order of post war international relations’.²⁸ Consequently, human rights have the power to limit State actions whether or not that State has human rights laws – this is what has made them so useful in relation to civil rights issues. Nevertheless, human rights have expanded beyond this orthodox remit and courts have used interpersonal interests or rights as justification in their judgements. Recent case law demonstrates this expansion.

In the most recent assisted dying case, judges have stated that ‘the involvement of a third party raises the problem of the effect on other vulnerable people, which the unaided suicide does not’²⁹ and assisted dying is ‘rationally connected to ... safeguarding the lives of those who are weak and vulnerable and may, without protection, feel pressured into agreeing to die’.³⁰ More importantly, the court explicitly recognised that this imposed constraints on the right-holder. ‘It can fairly be said that in many cases this approach will deprive those closest to the patient of the means of enabling him to kill himself’.³¹ These statements suggest, at the very least, that the actions of third parties were a factor in the court’s reasoning – perhaps even a decisive factor.

These decisions mean that human rights cannot be used to authorise particular third-party actions through the power to decide when and how those rights are exercised. This is problematic because it means that rights cannot be alienated *even if* the right-holder desires to alienate a right or the exercise of a right when doing so will

²⁶ Sieghart (n1) 8.

²⁷ Ibid 15.

²⁸ Michael Ignatieff, *Human Rights as Politics and Idolatry* (Princeton University Press 2011) 5.

²⁹ *Nicklinson* (n16) SC Lord Sumption JSC para.215.

³⁰ Ibid Lord Dyson MR and Elias LJ para.36.

³¹ Ibid Lord Sumption para.252.

better fulfil their preferences than not being able to do so. As human rights apply to a heterogeneous population of agents with different and competing – perhaps equally (in)valid – moral systems, constraint of this kind cannot be justified on the basis that *others* think it is wrong to permit assisted dying (or that is wrong to be act as a surrogate). Furthermore, if there is no risk to other agents from permitting the waiver of the exercise of a right, this constraint is not needed to ensure co-operation because the system still protects other agents from abuse.

If it becomes possible to alienate the protections of a right, judges must ‘ascertain whether some human being ... has retained them or whether they have been alienated’ thus ensuring that checks remained in place when it is unclear whether an agent has waived the protections of their rights.³² Agents would still have a reason to comply with those constraints that human rights impose upon them to protect other agents because this is necessary for maintaining co-operation. For example, it would be rational for a surrogate mother to be able to waive her status as the legal mother³³ making the process of surrogacy and parenthood much clearer whilst protecting those who do not want to forgo the exercise of their rights from being forced to do so. If the potential father in *Evans* had been able to alienate his rights with regard to the child – forgoing his right of access and responsibility of maintenance – this may have changed his view about the creation of a child with the fertilised egg. Unfortunately, the judicial interpretation in that case precluded such an option from being available.

As a result of this some agents have sought to act outside English jurisdiction. For example, clinics such as Dignitas³⁴ in Switzerland, cases such as *W v H (Child Abduction: Surrogacy) (No1)*³⁵ (which concerned a Californian surrogacy agreement with a British surrogate) and *Re X (A child) (Surrogacy: Time Limit)*³⁶ (which involved an international surrogate in India) are all cases where the agents sought to act outside English law because judges do not recognise the waiving of the exercise of rights. This is particularly complicated in surrogacy because surrogacy arrangements are not legally enforceable³⁷ in English law and a subsequent decision, made on the

³² Veer (n7) 174.

³³ Human Fertilisation and Embryology Act 1990 s.27(1) and Human Fertilisation and Embryology Act 2008 s.33(1).

³⁴ See http://www.dignitas.ch/index.php?option=com_content&view=article&id=22&lang=en accessed 14/12/2014.

³⁵ [2002] 1 F.L.R. 1008.

³⁶ [2014] EWHC 3135 (Fam).

³⁷ Surrogacy Arrangements Act 1985 s.1A: ‘No surrogacy arrangement is enforceable by or against any of the persons making it’.

basis of child welfare, can render the arrangement void. In *JP v LP SP CP*³⁸, the surrogate, who as the birth mother is the legal mother until a parental order is made, was placed in the invidious position of retaining legal parenthood when the commissioning couple's relationship broke down and the commissioning mother failed to submit the parental order request.

If the surrogate had been able to forgo her legal status as the mother prior to birth, perhaps through the surrogacy arrangement itself, such a situation would not have occurred because the legal mother would have always been the commissioning woman.³⁹ The alternative mechanism of adoption also permits an agent to forgo their status as a parent but this post-natal process may be less efficient than pre-natal waiver because – as with parental orders – this requires an application after the surrogate has acquired the status of legal mother with the result that there is a time lag between the child's birth and the parental status of the commissioning parent(s). Moreover, in adoption it is uncertain who will end up with the child while in surrogacy case this uncertainty is much reduced. This creates a disjunction between the offspring and their 'parents' (i.e. the surrogate and the commissioning agent or agents) which opens up the possibility of conflict over parental status. If the legal status of each participant could be defined prior to the offspring's birth this would eliminate that time lag removing the possibility of anyone other than the commissioning agent(s) being the parent(s).

Moreover, in assisted dying cases (limited to those involving mentally competent agents who want to end their life) agents are prevented from ending their lives because their right to life (ECHR Art.2) does not entail a right to receive active assistance at the end of life.⁴⁰ In this situation someone wants to extinguish or forgo the protections afforded by their right by removing the duty on others not to kill them but are prohibited on the basis of that *very* right. A number of cases brought before the European Court of Human Rights (hereafter the 'ECtHR') confirms the problems of this position. The justices of the ECtHR in *Pretty v United Kingdom*⁴¹ held that the

³⁸ [2014] EWHC 595 (Fam).

³⁹ As it was the surrogate was 'prohibited from exercising any parental responsibility for CP without the leave of the court' *JP v LP SP CP* para.36.

⁴⁰ Unlike some other rights which entail having to use the right so as *not* to realise the object of the right, for example the right to found a family includes the right to not have a family. See: *Evans v United Kingdom* (n19).

⁴¹ [2002] 35 E.H.R.R. 1.

‘Court is not persuaded that ‘the right to life’ guaranteed in Article 2 can be interpreted as involving a negative aspect’.⁴² In other words:

Article 2 cannot, without a distortion of language, be interpreted as conferring the diametrically opposite right, namely a right to die, nor can it create a right to self-determination in the sense of conferring on an individual the entitlement to choose death rather than life.⁴³

Another line of argument presented to the Court of Appeal, prior to the case reaching the ECtHR, was also rejected. Counsel for Pretty argued that ‘the proposed undertaking would *not be incompatible* with Article 2. This article protects, not life itself, but the individual's right to life from attacks by third parties, including the State’.⁴⁴ The court agreed with this argument to the extent that the right to life ‘does not *require* the State to take positive steps to force life upon the unwilling’ but ‘[f]or a third person to take active steps deliberately to deprive another of life, even with the consent of the person thus deprived, is forbidden by the Article’.⁴⁵ In *Nicklinson*, Lord Sumption JSC stated that a person who wants to die ‘does not have a right to call on a third party to help him to end his life’.⁴⁶

Here, we can see clearly how this results in constraint upon the right-holder themselves by limiting what others can do to the right-holder even if the right-holder wants to authorise that action by forgoing the constraint their right imposes on others. It cannot be emphasised enough that construing the right to life as prohibiting third party assistance is *not* forbidden by the article but forbidden by the interpretation of the article judicial officials have adopted. The judges could have constructed the right to life as containing the power to voluntary consent to forgo the right’s protections. This is important because it highlights that a different interpretation of human rights is possible and may be desirable because it frees agents from the unjustified imposition of constraint on the basis of moral concerns or because of the unsubstantiated risk to others.

⁴² Ibid 2.

⁴³ Ibid 2.

⁴⁴ *The Queen on the Application of Dianne Pretty v Director of Public Prosecutions* [2001] WL 1171775 (Queen’s Bench) para.40.

⁴⁵ Ibid para.41.

⁴⁶ *Nicklinson* (n16) SC para.255.

The moral infection of human rights law is clearest in the application of the moral concept of the sanctity of life. The House of Lords stated that ‘the language of Article 2 reflected the sanctity of life’ and that the right to life ‘could not be interpreted as conferring a right to self-determination in relation to life and death and assistance in choosing death’.⁴⁷ The Court of Appeal in the recent case brought by Tony Nicklinson stated that ‘[t]he sanctity of life was a more fundamental principle of the common law. There was no right to commit suicide’.⁴⁸ This was echoed in the Supreme Court by Lord Dyson MR and Elias LJ, who put it this way: ‘if there is no right to kill yourself, there can be no right, fundamental or otherwise, to require the state to allow others to assist you to die or to kill you’.⁴⁹ Again, we can see that the courts reject either a separate right to die, a right to commit suicide, a right to receive assistance or a right to give up one’s right or the duties imposed upon others.

In contrast, Lord Kerr of Tongahmore JSC and Baroness Hale of Richmond JSC dissented from the majority, arguing that the court should have made a declaration of incompatibility, albeit on the basis of Article 8 (right private and family life) not Article 2 (right life). Baroness Hale suggests that English law is incompatible with the European Convention ‘[n]ot because it contains a general prohibition on assisting or encouraging suicide, but because it fails to admit of any exceptions’.⁵⁰ In *Hass v Switzerland*⁵¹ the court concluded that ‘an individual’s right to decide by what means and at what point his or her life will end ... is one of the aspects of the right to respect for private life within the meaning of Article 8 of the Convention’.⁵² Lord Kerr argued, on the basis of Article 8, that ‘[t]here is no question of the claimants claiming that they should be assisted by the state to do what they want to do’⁵³ rather it should ‘merely tolerate it’.⁵⁴ Baroness Hale accepted that protecting the lives of others ‘is certainly enough to justify a general ban on assisting suicide. But it is difficult to accept that it is sufficient to justify a universal ban.’⁵⁵

⁴⁷ *Regina (Pretty) v Director of Public Prosecutions* [2001] UKHL 61, 61.

⁴⁸ *R. (on the application of Nicklinson) v Ministry of Justice* [2013] EWCA Civ 961 (Appeal Court) 874: despite suicide no longer being illegal or a criminal offence.

⁴⁹ *Nicklinson* (n16) SC para.55.

⁵⁰ *Ibid* para.301.

⁵¹ (2011) 53 EHRR 1169.

⁵² *Ibid* para.51.

⁵³ *Nicklinson* (n16) SC Lord Kerr para.329.

⁵⁴ *Ibid* SC Lord Kerr, para.331.

⁵⁵ *Ibid* SC Baroness Hale para.313.

Moreover, the judiciary do recognise that parental status can be optional. The provision of a ‘formal set of “agreed fatherhood conditions”’⁵⁶ and the ‘mechanism for same-sex couples to have both parties registered as the legal parents’ demonstrates that the judiciary recognises – at least in some limited sense – that taking on parenthood and parental duties is optional.⁵⁷ Thus, the identification of those with parental responsibility is separable from biological connections and that parental status can be assigned by agreement and that parental rights can be revoked (as in *W v H (No1)*). The judicial system can therefore interpret human rights as containing a greater level of control than is currently recognised. There are, then, some senior members of the judiciary who *do* conceive of the protections or entitlements of human rights as something that the right-holder can forgo when they voluntarily decide to do so even if this is an exception rather than the rule.

Accordingly, ‘English law curtails a person’s right to bodily autonomy in the interests of protecting that person’s life even against her own wishes’ and consequently a ‘person’s own wishes are therefore not determinate of what can or must be done’.⁵⁸ Again, there ‘is as yet nothing in Convention jurisprudence to suggest that this is a right which can be waived, rather the reverse’ – the constraints that are imposed on right-holders by their rights cannot be waived or forgone.⁵⁹ The position adopted by both the English courts and the European Court of Human Rights is consistent with the principle of the international system of human rights law that ‘no human being can be deprived of any of those rights, by the act of any ruler or *even by his own act*’.⁶⁰ But there is a distinction between the loss of the right itself and the loss of the legal protections that right creates –as we shall see later, this is an important distinction because it allows us to restructure rights to permit the waiver of protections without relinquishing the right itself.

But the constraints of the current regime are objectionable because it prevents agents from acting to fulfil their preferences when allowing them to do so does not jeopardise the protection of other agents or the co-operative scheme. Who is better placed to decide whether an agent should assume parental responsibility in a surrogacy arrangement? Who is better placed to decide whether death is the preferable

⁵⁶ Nils Hoppe and Jose Miola, *Medical Law and Medical Ethics* (Cambridge University Press 2014) 170. See: Human Fertilisation and Embryology Authority 2008 s.37.

⁵⁷ *Ibid* 171. See: Human Fertilisation and Embryology Authority 2008 s.44.

⁵⁸ *Pretty v DPP* QB (n44) para.37.

⁵⁹ *Ibid* para.41.

⁶⁰ Sieghart (n1) 8 (my emphasis).

option? How can preventing someone from achieving their most desired outcome be justified when this does not threaten others or the co-operative scheme? The answer is that only by moral imperialism can such constraint be justified by this is nothing more than the domination of a particular belief system for, as I argued in chapter one, moral claims are not truths about existence that apply to all agents.

This is not to say that the judiciary are not sympathetic to the argument made before them. Indeed the ‘courts have developed the law in this area with a sense of disconnect between what the law requires and what they want to do themselves’.⁶¹ End of life cases have consistently shown that the courts ‘are broadly sympathetic to the idea that adults who wish to die should be allowed to do so’ but they are limited by the legal context in which they operate – in England the rule of Parliamentary sovereignty plays a significant role in limiting what interpretations judges are prepared to deploy.⁶² As Toulson LJ notes:

A decision by the court to alter the common law so as to create a defence to murder in the case of active voluntary euthanasia would be to introduce a major change in an area where there are strongly held conflicting views, where Parliament has rejected attempts to introduce such a change, and where the result would be to create uncertainty rather than certainty. To do so would be to usurp the role of Parliament.⁶³

Even so, if Parliament is not prepared to deal with these issues (whether by referendum or a free House of Commons vote) then judges may eventually have to act, bringing about a change in the interpretation of human rights law and common law (for crimes like murder) – which, ironically, is where judges have traditionally had most leeway in creating law. The recent Assisted Dying (No.2) Bill failed to pass its second reading on 11 September 2015, consequently judges may feel that they cannot change the law after such recent legislative action – this does not of course affect whether a system of human rights with a waiver is a more rational system than the current one.

⁶¹ Hoppe and Miola (n51) 281.

⁶² Ibid 281.

⁶³ *Nicklinson* (n16) SC Toulson LJ, para.84.

These problems arise because right-holders cannot waive the exercise of their rights or forgo their rights and this concept is enshrined in the notion of inalienability – but inalienability can mean either the right is inalienable or the good that that right secures is inalienable. This is conflated by judges – notice that in the above quotations judges switch between talking about deprivation of life, the right to life and a right to end life without being clear whether it is the right, the good it secures or both which is inalienable. The theory of rational constraint, which treats legal rules as second-order constraints, provides the justification for refining inalienability to remove this confusion and propose an alternate construction of human rights that reduces an agent’s constraint without weakening the duties imposed on others. We now need to consider what conceptual tools can bring this about and explain how this system would function.

The Notion of Inalienability

The orthodox conception of alienating arises due to the ‘failure to distinguish alienating from ... forfeiting and annulling, and also a failure to distinguish between two possible interpretations of alienating, namely waiving and relinquishing’.⁶⁴ First, we can distinguish between forfeiting and alienating as follows: ‘The distinction is roughly between losing a right through one’s fault or error, on the one hand, and voluntarily giving the right away, on the other’.⁶⁵ Forfeiting does not require that the right-holder desires the loss of their right for ‘[a]s soon as one’s conduct falls below the qualifying standards one loses the right, whether one likes it or not’.⁶⁶ Thus:

*A nonforfeitable right is one that a person cannot lose through his own blundering or wrongdoing; an inalienable right is one that a person cannot give away or dispense with through his own deliberate choice.*⁶⁷

Thus, legal human rights are not forfeitable because one’s conduct does not have any relevance to species membership and they are explicitly designed to be exempt from any ‘performance’ criteria in order to prevent States revoking an individual’s rights.

⁶⁴ Joel Feinberg, ‘Voluntary Euthanasia and the Inalienable Right to Life’ (1978) 7 *Philosophy and Public Affairs* 93, 111.

⁶⁵ *Ibid* 111.

⁶⁶ *Ibid* 111.

⁶⁷ *Ibid* 112.

The inalienable possession of a right is distinct from the exceptions that permit the *lawful* infringement of a right but only to the degree needed to achieve a legitimate aim.⁶⁸ There is no comparable power held by right-holders, only States can claim an exception and thus lawfully infringe a right.

The second distinction is between alienating and annulling. The ‘major source of confusion in criticisms of the doctrine of inalienable rights ... might have been obviated, as B.A. Richards suggests, by consulting a good dictionary’.⁶⁹ Richards’s review of Webster’s definitions of the terms inalienable and inalienable is illuminating and worth the following lengthy quotations:

The term ‘alienate’ itself has various meanings, and two of them are worth noticing. One is as follows: to convey or transfer to another, as title, property, right; to part voluntarily with the ownership of. The other is: to cause to be transferred or withdrawn.⁷⁰

And:

‘Also of interest is the note appended to the definition of the word “inalienable”. It identifies “indefeasible” as a synonym and then proceeds to distinguish the meanings of the two words as follows: indefeasible, what one cannot be deprived of without one’s consent; inalienable, what one cannot give away or dispose of even if one wishes.’⁷¹

Richards demonstrates that inalienable *can* be interpreted as containing an element of *voluntarily forgoing* one’s rights without permitting the removal of that right by others. If one reads ‘an “inalienable right” [as] one that could not be cancelled or withdrawn by the state’⁷² then ‘the word “indefeasible” for a right that cannot be taken away from its possessor by others’ may be more appropriate.⁷³ At the very least, inalienable does not *necessarily* mean that the right-holder cannot give up their

⁶⁸ Craig and de Burca (n2) 526.

⁶⁹ Feinberg (n64) 112.

⁷⁰ B.A. Richards, ‘Inalienable Rights: Recent Criticism and Old Doctrine’ (1969) 29 *Philosophy and Phenomenological Research* 391, 398 fn.31.

⁷¹ *Ibid* 398 fn.31.

⁷² Feinberg (n64) 112-3.

⁷³ *Ibid* 113.

right(s) in some way. More positively, it suggests that we can possess non-forfeitable indefeasible rights that give the right-holder more choice over their actions than inalienable rights.

The former interpretation of inalienability as not revocable by the State or the right-holder (as opposed to the exercise of a right) is the prevailing interpretation of inalienable human rights for human rights ‘are not acquired nor can they be transferred, disposed of or extinguished, by any act or event: according to the classical theory, now reflected in the international standards, they are “inalienable”’.⁷⁴ The key phrase here is *by any act or event* which would seem to preclude both alienating in the sense of giving up one’s right *and* alienating in the sense of being taken away. This is reinforced by that statement that ‘[f]or a third person to take active steps deliberately to deprive another of life, *even with the consent of the person thus deprived*, is forbidden by the Article’ and the other judicial comments quoted above.⁷⁵

Having distinguished the variety of terms that can be used to describe rights, we can now turn to final most important distinction between waiving and relinquishing. The distinction is ‘between waiving exercise of a right that one continues to possess and relinquishing one’s very possession of the right’.⁷⁶ Alternatively, ‘[w]hat exactly is it that cannot be alienated when one has an inalienable right to X – X itself or the right to X’?⁷⁷ How inalienability is constructed will determine whether one can waive the exercise of a right in particular circumstances whilst retaining that right. To waive a right is to ‘exercise my power to release others from correlative duties to me, to desist from claiming my right against them’.⁷⁸ This would allow us to release others from the duties they owe us and that we owe others – medical professionals could be released from the duty not to kill, surrogates could be released from the duty of parenthood, separated couples could mutually release each other from duties owed as parents.

An example will add lucidity to this distinction. Consider the right to property: if ‘I give all my property away, I have not abandoned the discretionary right to acquire (or re-acquire) property; rather I have chosen to exercise that right in a

⁷⁴ Sieghart (n1) 17.

⁷⁵ *R(Pretty) v DPP* QB (n44) para.41 (my emphasis).

⁷⁶ Feinberg (n64) 114.

⁷⁷ *Ibid* 114.

⁷⁸ *Ibid* 114.

particular, eccentric, way'.⁷⁹ This is an example of waiving the good that a right allows me to possess without relinquishing the right to property itself. Now 'imagine a constitutional order and a legal system in which the right to property itself is alienable'.⁸⁰ This is an example of relinquishing the right to X and thus the possession of X – '[i]f one were thus permitted to relinquish the right [to property] permanently, one could possess objects and occupy places but never *own* them'.⁸¹ An historical example might be joining a monastery and giving up all rights to property, retaining only the use and possession of those owned collectively by the monastery.

We can even imagine a scenario in which this distinction applies to the right to life, albeit a rather bizarre scenario. Imagine 'a legal system so permissive that it allows one to formally contract with another ... to put one's life – one's continued existence – in the other's legal power'.⁸² One would continue to possess all other rights (e.g. to property, to vote, to a private and family life and so on) while no longer possessing the right to life, which is now held by another, who can end one's life when they so choose. Yet one cannot temporarily give up the good that the right to life secures because 'I cannot destroy my life for a period of time while maintaining my discretionary right to re-acquire life'.⁸³ Consequently, waiving the exercise of the right to life means that one ceases to be a right-holder. Although extreme, this example shows the conceptual distinction between the right to a good and the good itself but it also shows that not all rights can operate in the same way because of the goods they involve. Hence, waiver of the right to found a family might only be permitted prior to the birth of any offspring to prevent parents reneging on their responsibilities after the child comes to exist and thus transfer the costs of parenthood onto others.

Parental Responsibility and Limited Waiver

It might be argued that potential parents may use their waiver to evade responsibilities to offspring due to a change of mind or when the offspring is unplanned as a consequence of unprotected sexual activity or from an accident (such as contraceptive failure). This is problematic because of the issue of child support or maintenance, the

⁷⁹ Ibid 115.

⁸⁰ Ibid 115.

⁸¹ Ibid 115.

⁸² Ibid 117.

⁸³ Ibid 117.

possibility of all parties waiving their parental claims and duties and the disproportionate effect on women. These problems are not as great as they first appear but only by carefully circumscribing the scope of the waiver can the issues become manageable.

By allowing the waiver of rights relating to reproduction there is a risk of children being left orphaned and/or without any financial support beyond that provided by the State. The existence of the power to waive parental status may increase the probability of children being rendered parentless if the reproductive arrangement breaks down or the sole parent dies. We might counter this by pointing out that parentless children could be included in the current system that deals with orphans or those given up for adoption through the social services system. But an increase in children in care may be undesirable because of the costs it imposes on public finances – indeed, waiving one’s parental status transfers the costs (and duties) to other agents and this might justify assigning parenthood to *someone* if agents cannot decide amongst themselves who that someone is. To minimise the occurrence of this outcome we should restrict the option of waiving parental status to pre-conception circumstances in order to prevent a prospective parent from changing their mind during the gestational period – *pre-conception* is a suitable point at which limit the waiver of parental status because the prospective parents will start expending resources in expectation of a child and because it covers the possibility of pregnancy through unprotected sex or failure of a contraceptive.

This circumscription would be further justified by the relationship between the procreating agents which raises the problem of the disproportionate effect upon women. It is easier for the non-gestational parent (whether male or female) to withdraw from the procreative process, with the risk that *women* will suffer disproportionately as the individual carrying the offspring. It is easier to assign women as legal parents – as the law currently does – by treating the birth mother as the legal mother but this relatively straightforward criterion is not so easily applied to men or non-gestational mothers. This is why waiver should require an agreement *prior to conception*, not because this will affect the child but because it will affect the interaction between the prospective parents and the surrogate. Applying this pre-conception restriction would be easier in artificial reproduction than in coital reproduction due to the greater evidential resources and separable technological

processes of the artificial reproduction. Coital reproducers may have to submit a default system which automatically assigns parenthood to agents.

It seems clear that those with parental status should not be able to waive their rights *after* birth because this would render the child parentless – at least without the compensation of putting the child into a transfer process (such as adoption or parental orders) in which another agent voluntarily takes on the role of parent. The purpose of limiting the use of a waiver is to prevent unilateral transfers of the costs of raising children to the State and to prevent one parent from withdrawing from the joint enterprise of founding a family (if there is more than one prospective parent). On this basis we might restrict waiver to non-coitally reproducing parents who have the time and conscious forethought to establish their parental status and to anticipate changes in mind or circumstance. This does not negate the usefulness of a waiver in helping to increase an agent's ability to fulfil their preferences or the clarity that can come from being able to determine parental status prior to birth.

Consequently, it may be best to retain a default system of parenthood under which an agent can waive their parental status only in certain circumstances – we might restrict it specifically to pre-conception, artificial reproduction, explicit surrogacy arrangements, to all three, to all artificial procreative acts or all procreation. However, all those who reproduce without assistance would be subject to the default system (which would be the one that currently operates) because they assume a risk of pregnancy that those seeking assistance do not and parental status a consequence of this risk rather than a goal. But the point remains that none of this provides a reason for prohibiting agents from assigning parental rights and responsibilities themselves, it merely suggests that a default system is needed to reduce abandonment by prospective parents changing their mind and careful limitations on the exercise of the waiver is needed to prevent the unilateral transfer of costs to others.

The problem that waiver generates concerns how agents assign parental responsibility between themselves – a voluntary system is preferable (because of the reduced costs) to one that assigns parental status contrary to wishes of agents. But not all agents will comply with their agreements become a parent – as we saw in chapter two compliance is a significant problem in interaction and as the case of *Evans v United Kingdom* showed circumstances do change. In these circumstances imposing parental status on, say, the commissioning couple (even if they no longer want the child) may be justified because they are seeking to transfer the costs of their decision

to others – given that these costs only arise because of them other agents may rationally prevent such a transfer occurring.

In relation to the right to life a limited waiver may also be the best form of waiver – for example, by restricting it to those who are terminally ill. Again this may be justified on the effect on others due to the responsibilities imposed on others in administering the estate. In cases of terminal illness the agents action does not generate new costs but merely affect the timing of those costs. However, those who are not terminally ill but who wish to die may require third-party intervention and it is unclear whether to permit this or not because the costs to third-parties are much smaller than the benefit to the agent. This is unlike parental duties which can impose substantial costs on other agents while the parents are still around and where parenthood is to a great extent optional (unlike death).

For these reasons the waiver system should be limited to circumstances prior to the existence of any child to prevent parents unilaterally transferring their costs to others, reduce the incidence of parentless children and to prevent agents avoiding their duties when they agree to have offspring but change their mind at a later point. A limited waiver for the right to life may be justified restricting it to those diagnosed with a terminal illness but it is not clear that it could be so restricted when the benefit to the agent is greater than the costs their death would imposed. Nonetheless, a limited-waiver does not negate the benefits it would confer on agents by being able to self-determine and establish their parental status prior to the existence of any offspring. So in what circumstances would this waiver be useful? The answer is that it will be most useful in cases of artificial reproduction – whether this is surrogacy, artificial insemination, in vitro fertilisation or any other technological means of reproduction which require conscious choices on the part of the prospective parents. I will now focus on reproductive cases because this will help outline how the most limited form of waiver is still beneficial for agents.

The Application of Waivable Human Rights

The following cases serve an illustrative function demonstrating that there are situations in which it may be rational to waive the exercise of a right in relation to reproduction. Although these are not all human rights cases they are relevant because if an agent waives their domestic legal right but retain their human right they would always have the option of enforcing their rights – at least in European jurisdictions

that permit applications to the European Court of Human Rights – and if waiver can be shown to work with human rights, it can easily be applied to standard legal rights.

As the right to marry and found a family allows one to either found a family or not, it would seem that the power to waive the exercise of the right would be redundant. However, reflection on the case of *Ferguson v McKiernan*⁸⁴ from the United States of America shows that it might be desirable for one to waive the good that the right secures. ‘Ferguson and McKiernan orally agreed he [McKiernan] would provide the necessary sperm for IVF procedures in return for Ferguson's promise not to seek financial support for the children’.⁸⁵ Here we have a case of an agent, McKiernan, who wants to waive his *claims* to legal parenthood in order to or on the condition of forgoing the *duties* that it would impose. Note once again, the distinction between waiving the right and waiving what the right protects – McKiernan wanted to waive his right in relation to a particular act of reproduction but he did not want to relinquish his right to found a family itself.

In this case, the US courts determined that this arrangement was equivalent to standard sperm donation through a clinic, particularly as the father had not had any contact with the mother or the children (twins) for the first five years until she brought proceedings against him for financial support. The court of first instance would have concluded that they had ‘formed a legally binding oral agreement in which McKiernan relinquished any and all parental rights and obligations to the resultant child(ren) in return for Ferguson's promise not to seek child support’⁸⁶ but the court found the agreement unenforceable. The Supreme Court of Pennsylvania ultimately reversed the decision of the court of first instance declaring the verbal agreement an enforceable contract. While this case concerns US law, it is instructive for it demonstrates the idea of waiving one’s rights in family matters can lead to a greater degree of preference fulfilment for an agent than the simple restrictive interpretation inalienability and it can prevent post-hoc impositions of duties.

In English law, the case *W v H (No1)* reached a decision that extinguished the surrogates rights over the children due to an ‘order declaring that W and B [the commissioning couple] should have custody of the children at birth, and that the

⁸⁴ 855 A.2d 121 (Pa. Super 2004).

⁸⁵ David Rohwedder, ‘*Ferguson v. McKiernan*: Can a Sperm Donor Be Held Liable for Child Support After the Recipient Has Contractually Waived that Right?’ (2008) 32 *American Journal of Trial Advocacy* 229, 230.

⁸⁶ *Ibid* 231.

surrogate mother did not have any parental responsibility or rights'.⁸⁷ Although this was based on the welfare of the children this highlights the potential usefulness of being able to waive a right – if surrogacy arrangements were legally enforceable and permitted the surrogate to waive the exercise of her right to private and family life and her right to found a family the commissioning couple would be *from birth* the legal parents, not the surrogate. Thus the use of waiver can make the familial status of each participating agent clear from the very beginning minimising the costs associated with surrogacy and other forms of artificial reproduction – particularly when this necessitates judicial intervention.

As previously noted surrogacy arrangements are not legally enforceable under English law⁸⁸ – although surrogacy arrangements can be recognised by the judiciary but only on the basis of the welfare of the child – which leads to some significant problems. In *W v H (No 1)* the surrogate attempted to keep the child and as she was the legal mother it was more difficult for the commissioning couple to gain access to the child – this may also explain why some use partial surrogacy⁸⁹ which ensures a genetic connection between the offspring and the commissioning male. In *JP v LP, SP, CP* the surrogacy arrangement was made illegally as it was drawn up as a commercial contract which is a criminal offence⁹⁰ and the legal parentage of the child was thrown into doubt by the breakdown of the commissioning couple's relationship. In *Re P*, the surrogate mother kept the child because 'as in all wardship disputes, the welfare of the children was the paramount consideration' and the court found in her favour.⁹¹ The 'surrogacy agreement was relevant only if it reflected adversely on the fitness of the potential custodians to look after the children'.⁹² Under the waiver proposed here the welfare of the child may not be the determining characteristic because the child could be placed with its parents immediately and the agreement between the agents would be legally enforceable (if it were not enforceable this would

⁸⁷ J. Kenyon Mason and Graeme Laurie, *Law and Medical Ethics* (Oxford University Press 2006) 287.

⁸⁸ Surrogacy Arrangements Act 1985 s.1A.

⁸⁹ Margot Brazier and Emma Cave, *Medicine, Patients and the Law* (5th ed, Penguin Books 2011) 375.

⁹⁰ Surrogacy Arrangements Act 1985 s.2. Mrs Justice King highlighted this stating: 'The parties agreed and an agreement was prepared by a firm of Birmingham solicitors. The solicitors were in fact committing a criminal offence as, whilst such agreements can lawfully be drawn up free of charge, the solicitors in preparing and charging for the preparation of the agreement were *negotiating surrogacy arrangements on a commercial basis*' para.7 (original emphasis).

⁹¹ *Re P (Minors) (Wardship: Surrogacy)* [1987] 2 F.L.R. 421 Case Analysis 1.

⁹² *Ibid* 1.

replicate the problems with the current regime of unenforceable surrogacy arrangements).

If the exercise of rights could be waived this would allow the agents involved the self-determination to establish their parental statuses prior to the birth of any offspring – this is especially useful for surrogates and informal⁹³ sperm donors (like McKiernan) who do not want to have any parental duties to the offspring. It would also enable the parties to develop contingent parenthood – perhaps, the surrogate agrees only to the waiver of their right if the commissioning couple are alive and able to care for the child (they might be seriously ill, in prison or otherwise unable to care for the child) and if they are not, then the surrogate accepts legal parenthood. Such a system would give the parties to the arrangement much more flexibility, control and clarity over the exact relationship between each other and to the offspring.

The argument here is that agents should be able to change their legal status as parents from the default system – whether the default system needs reformation or not is a separate question – because this leads to more beneficial outcomes by allowing the greater fulfilment of preferences, reducing litigation and introducing certainty into the reproductive process. Note that surrogates or sperm donors would not be relinquishing their rights because they would still be able to exercise their rights in other instances of reproduction – a sperm donor does not forfeit their right to have children at another point in time.

But such a proposal raises the ‘very real question of whether we can or should, from a legal viewpoint, compel the surrogate to actually give up the baby’⁹⁴ – if commercial surrogacy contracts were legal this would raise the question of specific performance or compensation and if the surrogate was not the legal parent then retaining possession of the offspring might amount to child abduction. It may be that the process of surrogacy becomes a much longer or reflective process to reduce the occurrence of these kinds of conflict but, although using the waiver the frequency of these situations would be reduced but not eliminated, some judicial intervention will be required – unfortunately, this may be inevitable.

⁹³ That is a sperm donor who does not go through a clinic but who provides sperm to someone they know.

⁹⁴ Hoppe and Miola (n51) 181.

The limitation of waiver is a response to ‘the very complex issue of consent up to the very point of implantation to contend with’.⁹⁵ In all the above cases the judiciary seem to demand some pre-conception arrangement and, if this is the correct reading, the theoretical problems does not materialise *as long as* this waiver occurs before (in the case of artificial reproduction) implantation of any embryos. Deciding prior to conception is important – not because it affects the child who will result – but because it establishes the relationship between those involved in procreation and their legal status. This would have useful practical consequences.

In the conjoined cases of *Evans v Amicus Healthcare Ltd* and *Hadley v Midland Fertility Service Ltd and others*⁹⁶ it was suggested that the male partner may have consented if he would not have been subject to parental and financial duties. Ms Evans submitted that ‘there was an agreement between them that she could use the embryos provided he was not named on the birth certificate as the father and would not be financially liable for any maintenance’.⁹⁷ In the second case, Mr and Mrs Hadley had ‘discussions between them about Mrs Hadley being able to have the remaining embryos transferred into her on the basis that a resulting pregnancy would have no financial or other consequences for Mr Hadley’.⁹⁸ On the facts of the case Wall J rejected the existence of such any agreement but the important point is that Wall J would have concluded that “no such agreement would, in my judgment, have been binding”⁹⁹. If such an agreement had been binding it would have allowed the women to have the children they desire without imposing duties on the father.

This demonstrates that waiver in relation to family life is desirable for agents because it could resolve some of the difficulties that occur in some situations of artificial reproduction. But for this waiver to make any sense it must include waiver of one’s *duties*, as well as one’s *claims*, otherwise there would not be any point in waiving the benefits of the right to family life. The final part of this chapter focuses on explaining how the theoretical mechanism of waiver might work as a component of legal human rights.

⁹⁵ Ibid 176.

⁹⁶ [2003] EWHC 2161 (Fam).

⁹⁷ *Evans v Amicus Healthcare Limited* [2004] 2 W.L.R. 713, para.89.

⁹⁸ Ibid para.97.

⁹⁹ *Evans and Hadley* (n96) 740.

A Hohfeldian Analysis of Rights

Wesley Hohfeld's analysis of legal rights explains how this revised structure of legal human rights can function because it allows us to explain the main features of rights – that a right is possessed by someone, that it holds against others, that it can usually be exercised, waived, forfeited or renounced'.¹⁰⁰ A Hohfeldian analysis is useful for a number of reasons. First, the 'content of legal rights is typically defined with much more precision than the content of alleged nonlegal rights'.¹⁰¹ Hohfeld's analysis achieves (or moves towards achieving) this precision by distinguishing between the various meanings of 'right' – he is critical of the use of the term 'right' as it is used 'indiscriminately to cover what in a given case may be a privilege, a power, or an immunity'.¹⁰² The 'remedy for this conceptual muddle [between rights and other concepts] was to map the logical relations among a set of "fundamental conceptions" so as ... to distinguish rights from other items with which they were commonly confused'.¹⁰³ This is useful because:

Only when paradigm instances of rights are characterised in determinate language can they serve to distinguish between accurate and misleading general conceptions of rights. Nothing like this exists in the case of ethical rights, or even with the conventional rights of positive morality.¹⁰⁴

Consequently 'Hohfeld's fundamental legal conceptions remove several persistent ambiguities in the language of rights'¹⁰⁵ and he 'provide[s] a useful vocabulary for the analysis of complex legal positions'¹⁰⁶ which 'make[s] explicit the practical relevance of legal rights'.¹⁰⁷ Accordingly, a Hohfeldian analysis provides us with the tools to map the underlying structures of rights and to explain how these structures can be altered to permit waiver of the exercise of a human right.

¹⁰⁰ Carl Wellman, *A Theory of Rights* (Rowman & Allanheld 1985) 1.

¹⁰¹ *Ibid* 6.

¹⁰² Wesley Hohfeld 'Some Fundamental Legal Conceptions as Applied to Judicial Reasoning' (1913) 23 *The Yale Law Journal* 16, 30.

¹⁰³ Leonard Sumner, *The Moral Foundation of Rights* (Clarendon Press 1987) 18.

¹⁰⁴ Wellman (n 100) 6.

¹⁰⁵ *Ibid* 7.

¹⁰⁶ *Ibid* 10.

¹⁰⁷ *Ibid* 14.

Yet, Hohfeld's analysis is not without problems, '[t]o the philosopher's eye the most obvious [problem] ... is Hohfeld's failure to analyse any of his conceptions'.¹⁰⁸ Leonard Sumner argues that Hohfeld was mistaken to consider his conceptions as fundamental because 'it is possible to construct all of Hohfeld's conceptions out of two kinds of modality'.¹⁰⁹ This is understandable given that 'Hohfeld seems to have thought of the practical significance of legal terminology exclusively in terms of its application to courts cases'.¹¹⁰ Regardless, Hohfeld's analysis is useful for explanatory purposes as it stands so we will not interrogate Hohfeld's fundamental conceptions, although Sumner's first- and second-order modalities have an important contribution to make as well. Hohfeld's analysis identifies rights as consisting of the following relational concepts: (Opposites) rights/no-rights, privilege/duty, power/disability and immunity/liability; (Correlatives) right/duty, privilege/no-right, power/liability and immunity/disability. I shall focus on the conceptions of claims, duties and powers.

At this point we must note an important terminological difference. Hohfeld here uses the term 'right' to mean the relational correlative of a duty. Thus 'X has a right against Y that he shall stay off the former's land, the correlative (and equivalent) is that Y is under a duty toward X to stay off the place'.¹¹¹ In order to separate this stricter sense of the term *right* from its more general usage, 'the word "claim" would prove best' for the former.¹¹² This is useful because we 'need language that will distinguish between the complex right [the totality of relational positions that a right consists of] and the elements out of which it is constructed'.¹¹³ Following this *right* will refer to the totality of a set of Hohfeldian components and *claim* will mean the 'correlative of a relational duty'.¹¹⁴ This conception of claim entails that we 'cannot have claims *to do*, only claims *that others do*'.¹¹⁵

Paired up with the term 'duty' this provides the first-order concepts that make up the exercise of a right – but first duties must be defined. '[T]o have a duty under the law is necessarily to be duty-bound *to* some other party who has a correlative legal

¹⁰⁸ Sumner 19.

¹⁰⁹ Ibid 19.

¹¹⁰ Wellman (n100) 18.

¹¹¹ Hohfeld (n102) 32.

¹¹² Ibid 32.

¹¹³ Wellman (n100) 35.

¹¹⁴ Sumner (n103) 25.

¹¹⁵ Ibid 25.

claim holding against the duty-bearer'.¹¹⁶ We must note that 'we are not, however, committing ourselves to Hohfeld's assumption that all duties, or even all legal duties, are relational'.¹¹⁷ Not all definitions of a claim treat a duty as relational. For example, a compulsory right to vote imposes a duty to vote that is not owed to anyone. 'Hohfeld's conceptions are relations between two distinct parties' which contrasts with the 'simple deontic notion of a duty ... [which] is non-directional and thus non-relational'.¹¹⁸ Additionally, 'the actual system of international human rights law does include duties on states that are not formulated as correlatives of rights'.¹¹⁹ However, 'most legal duties are, as Hohfeld thought, duties to some second party' and, as we are concerned with the role of third-parties and human rights, the claim-duty relation is crucial to our understanding of how the waiver would work.¹²⁰

For present purposes the important fact is that agents cannot relieve third-parties of the duties their rights impose, while States can seek to reduce the duties they are under in particular cases through the use of exceptions to human rights, agents do not have an equivalent power. According to the theory of rational constraint duties can only be relational because constraint in relation to others is the only justification for limiting actions (whether that relation is between two specific individuals or between all agents is not, presently, as important). Notwithstanding this Buchanan argues for the possibility that the goals of human rights law may best be served by the imposition of mere duties without a corresponding right.¹²¹ This may be the case in relation to *State* obligations towards their citizens because there is not higher authority to whom States owe duties (although if the role of the United Nations expands this may change) and if they do not acknowledge duties to their citizen then they owe constraint to no one. The claim-duty relation controls the exercise of rights as we are concerned with them (because it imposes limits on the actions of third-parties) thus changing the way this relation works can remove those limits which would result in much more individual control over life.

¹¹⁶ Wellman (n100) 27.

¹¹⁷ Sumner (n103) 24-5.

¹¹⁸ Ibid 24.

¹¹⁹ Buchanan (n5) 40.

¹²⁰ Wellman (n100) 34.

¹²¹ Buchanan (n5) 132: 'those functions [of a human rights system] could be performed adequately by a system that does not feature individual rights, but relies only on (mere) duty norms'.

First- and Second-Order Components

First-order positions (i.e. the claim-duty relation) and second-order positions are the ‘two kinds of modalities’ to which Sumner referred earlier.¹²² The first-order concepts are based upon the ‘deontic modalities (required/forbidden/permitted)’ which provide the materials for constructing concepts such as a claim (permitted) and duty (required).¹²³ The capacity of second-order concepts are based upon the ‘alethic modalities; necessary, impossible and possible’, which modify the first-order claim-duty relation, provides a schema for the power of a right-holder to change when and how their rights operate.¹²⁴ The distinction between these modalities explains both the features of rights that are of concern here and how right-holders can modify the exercise of their rights in a particular instance without forfeiting them entirely.

These second-order concepts are as follows: a *power* allows one to alter normative relations with a correlative *liability* to having rules changed and an *immunity* against changes to normative relations with a correlative *disability* (*inability* may be a better more contemporary phrase) to alter those rules. Thus:

A change in a given legal relation may result (1) from some superadded fact or group of facts not under the volitional control of a human being (or human beings); or (2) from some superadded fact or group of facts which are under the volitional control of one or more human beings.¹²⁵

Accordingly, ‘I have a power to affect (that is to alter or sustain) some normative relation [the first-order position] just in case the rules of the system make it possible for me to do so’.¹²⁶ Those who are subject to the rules have a corresponding liability, an inability, to prevent the change being brought about by those rules. A legal power is ‘an ability, legal not mental or physical, of X to change some legal relation of Y by some voluntary act of bringing into existence one or more operative facts’.¹²⁷ For example, governments have the power to raise or lower taxes, reclassify who pays how much, include or exclude some agents from the tax system or abolish the tax

¹²² Sumner (n103) 19.

¹²³ Ibid 28.

¹²⁴ Ibid 28.

¹²⁵ Hohfeld (n102) 44.

¹²⁶ Sumner (n103) 28.

¹²⁷ Wellman (n100) 42.

system entirely while the citizens of that government has a liability to changes in the tax system.

In contrast, ‘I have an immunity against having some normative relation affected just in case the rules of the system make it impossible to affect the relation’.¹²⁸ Those who wish to change the rules protected by immunity have a disability, an inability to bring about that change. For example, Article 79(3) of the Basic Law of the Federal Republic of Germany makes certain changes to the constitution inadmissible thus creating a disability or inability to enact those proscribed changes. Immunity therefore imposes constraints on the power to alter legal relations or removes them entirely and creates a duty on others not to violate that immunity. We can see that immunities ‘are the second-order counterparts of claims’¹²⁹ and that disabilities (the restriction on conduct that immunities impose) are ‘the second-order counterparts of duties’¹³⁰. These second-order properties determine to what extent, if any, first-order rules can be altered¹³¹.

Thus a right-holder can possess the second-order power to alter the first-order duties that their right imposes on others without requiring a general change to their rights or the rights of others. If an agent can fulfil their preferences to a greater extent by altering the duties their rights impose, which would allow them to authorise third-party intervention, then a system permitting this is more rational than one without that power – particularly, as all of the protections provided by the power-less system are retained in the modified system. Consequently, it is legal powers of the second variety that provide the means by which a waiver can operate – those legal powers that are under the ‘volitional control of one or more human beings’.¹³² If an individual wants to end their life or forgo exercising their claims and the correlative duties imposed on others under the right to found a family, they must possess the second-order *power* to change the first-order positions of which their right consists. A surrogate would possess the power to change their claims and duties towards the offspring, a right-holder would be able to change the duties imposed upon medical personnel by waiving the duty to preserve life.

¹²⁸ Sumner (n103) 29.

¹²⁹ Ibid 31.

¹³⁰ Ibid 30.

¹³¹ This is similar to the amendment procedures of codified constitutions that determine which, to what extent and under what circumstances parts of the constitution can be altered.

¹³² Hohfeld (n102) 44.

Yet, the orthodox concept of inalienability prevents this possibility for a right is not under the control of any agent – inalienability operates more like immunity by prohibiting any change by any agent to the duties imposed by a claim. (Although, perversely, it allows the State to lawfully infringe those rights under certain circumstances.) Through parsing the concept of inalienability we saw that a distinction between rights as *indefeasible* and *non-forfeitable* to States and *waiving* the exercise of a right in a particular instance is conceptually sound. In order for the right-holder to be able to waive the exercise of their rights, they must possess the second-order *power* to change the operation of their right – specifically, by releasing others from the duties that human rights law imposes.

Conclusion: Refining Inalienability and the Legal Implications

The current orthodox conception of inalienability imposes greater constraint upon agents than is necessary because it does not allow the right-holder to alienate their rights in the sense of *waive*. If the mechanism of waiver were introduced, then, in right to life cases, a right-holder could remove the duty imposed upon medical staff *not to end life* permitting them to carry out active forms of euthanasia (assuming the medical practitioner is prepared to do or refers them to someone prepared to do so). Those who wish to die could then do so more quickly and with less distress. A surrogate could relinquish their status as legal mother much as those who put children up for adoption do. Parental status could be assigned prior to conception and/or implantation.

Agents should have this control because it will allow them to maximise their preferences in life *without* compromising the protections afforded to other agents. This waiver must be limited, however, to prevent the unilateral transfer of the costs of parenthood to other agents after the decision to become a parent has been made. Even so, a limited-waiver would extend the control agents have over their lives allowing them to fulfil their preferences to a greater degree. Consequently, preventing agents from exercising this control is unnecessary for maintaining the system of constraints generated by human law thus the constraint imposed upon right-holders is *unjustified* according the theory of rational constraint because it does not benefit other agents and reduces the benefit to those who would waive the exercise of their rights. What agents need is a second-order capacity to change the claim-duty relation allowing them greater control over when their rights impose duties on others.

Human rights law has developed from its historic constraining function on States extending to the imposition of constraints on individual right-holders by forbidding them the power to waive the exercise of their rights in particular instances when they wish to do so. Breaking the notion of inalienability down into the concepts of *indefeasible*, *nonforfeitable* and *waiver* would maximise the constraint on States by preventing them from infringing the area protected by the right whilst maximising the range of actions that right-holders can undertake. For the agent, possessing the legal power to waive the duties and/or claims of their rights would maximise their power relative to the State allowing them to fulfil more of their preferences while maintaining the duties towards other agents and the State.

This may not be the case for all human rights – for example, the prohibition on torture¹³³ might be rationally justified regardless of the circumstances because real torture (as opposed to role-play sadomasochism), by definition, causes involuntary suffering. In contrast, sadomasochistic sexual practices involve consensual activity undertaken for the benefit of the participants which distinguishes it from torture. The system of waiver may not, therefore, be applicable to all of one's human rights, although it is applicable to (at least) right to life, the right to found a family and the right to private and family life. Whether agents would be rational to possess a legal power to *relinquish* their rights is not as clear. If the power to waive the exercise of a right were present, it would seem to provide all the benefits of relinquishing without necessitating the loss of the right itself which would strip away *all* of the protections rights confer.

The key point is that a mechanism can be developed that grants greater control to agents over when and how they become parents, that clarifies the status of an agent before a child exists thus leading to a more efficient less costly system of reproduction which allows agents to establish and characterise a relationship of their choosing between themselves in reproductive arrangements. Such a system would give agent's the power to better fulfil their preferences in relation to parenthood which is a central, crucial and major factor in achieving the life an agent wants. For this reason the limited-waiver system is more rational than current legal orthodoxy and should be adopted in place of the contemporary human rights regime.

¹³³ ECHR Art.3: 'No one shall be subjected to torture or to inhuman or degrading treatment or punishment'.

Part 2

The following two chapters are based on two published papers. The thesis was originally conceived as an ‘alternative format’ thesis consisting of a number of published papers building on a common theoretical framework. During the later stages of writing it did, however become clear that adequately laying the theoretical foundations needed very considerable space, and that a thesis monograph was a better option. The two papers are thus much more topic focused than the rest of the thesis, and were written prior to the final explication of the theoretical framework.

Chapter Four

Potential persons and the welfare of the (potential) child test

This chapter has been published as ‘Potential persons and the welfare of the (potential) child test’ (2014) 14 *Medical Law International*.

Abstract

This article considers and critiques the theoretical basis of section 13(5) of the Human Fertilisation and Embryology Act 1990 (HFE Act)¹, which requires the welfare of the potential child to be taken into account when considering whether parents should be given access to fertility treatment. It will be argued that potential persons, that is, persons who do not yet exist, have no claims, interests or standing that can restrict the actions of actual persons. This claim will be based upon the necessity of existence before things can be said to affect a person. As persons are the subjects for whom ‘good’ and ‘bad’ outcomes apply, actions which establish the preconditions for their existence cannot be subjected to considerations of the effect on the potential person. This is because potential persons are not affected by actions, but are the consequence of actions. Prospective parents, for example, should not be prohibited from having disabled offspring on the basis of the effect on the child, as different decisions relating to that child will change which persons exist, and thus the necessary preconditions for value will change. Based upon this existential framework, it will be argued that only the interests of actual persons can constrain the actions of those involved in reproduction. Thus, the HFE Act’s formulation of the welfare test must be repealed

¹ As amended by the 2008 Act of the same name.

and only the interests of prospective parents and other actually existing people should be capable of constraining reproduction.

Keywords

Genetic, person, embryo, existence, welfare test

Introduction

Developments in fertility treatment and artificial reproductive technology have created an unprecedented level of control over the process of reproduction. Prospective parents now have many options for fulfilling their desire to have children; from artificial insemination to in vitro fertilization, from adoption to surrogacy and possibly through genetic modification. In response to the development of these technologies, the UK government established the Committee of Inquiry into Human Fertilisation and Embryology (1982–1984) to consider the implications and responses to these emerging reproductive technologies. The Committee produced a report (also known as the Warnock Report²), which then formed the basis for a draft bill that was passed into law as the Human Fertilisation and Embryology Act 1990 (hereafter the HFE Act). This Act created the Human Fertilisation and Embryology Authority, which has regulatory oversight of all fertility treatment and reproductive services (although this regulatory function may be transferred to the Care Quality Commission and others³). The HFE Act contains a provision in section 13(5), which requires that ‘the welfare of any child who may be born as a result of the treatment’ be taken into account.⁴ It is this ‘welfare of the child test’ as it exists in the HFE Act that is questioned and critiqued in this article.

It will be argued that the welfare test when applied to the potential child is problematic because it is too burdensome on prospective parents (hereafter parents) both because it requires that the parents are vetted to determine whether they are ‘suitable’ and because the test imposes obligations on the parents for the sake of the child (most noticeably in the case of disability). Moreover, if these obligations are complied with it may require the parents to abort the foetus or otherwise not carry

² Department of Health and Social Security, ‘Report of the Committee of Inquiry into Human Fertilisation and Embryology’, 1984.

³ Human Fertilisation and Embryology Authority, ‘Statement: HFEA Responds to Government’s Proposals on the Future of the HFEA’, 2012.

⁴ HFE Act 1990, section 13(5).

through the pregnancy, thus the child will never come to exist; and this would require that parents act for a child that will never exist. Equally, if genetic modification technology became available, it would require that parents use it to improve the welfare outcome of the potential child. The welfare of the potential child test therefore creates a discriminatory system in which those who cannot reproduce naturally are doubly penalized; they must prove their competence to be parents, and the test creates a strong prejudice against parents who want to produce a child who has, or might have, a disability. These restrictions thus need to be firmly grounded in order to justify the imposition of onerous obligations on parents and the denial of services to those who may not wish to undergo genetic testing for disability or those who may want to have a child with a particular disability.

The justification for restricting access to reproductive technology, as shown by section 13(5), is the ‘welfare of any [potential] child who may be born’.⁵ It is the benefits and welfare that accrue to the potential child that justify considering the conduct of the parents and the risk of disability. However, it will be argued that consideration of the welfare of the potential child is not valid because it is incorrect that a child can be benefitted or harmed by actions that cause them to exist; that is, actions which constitute the person cannot affect that same person because their existence is a necessary precondition for them to benefit or be harmed. As David Heyd states, ‘they [potential persons] do not meet certain preconditions of existence and identity’,⁶ which would allow them to be the object of beneficence or maleficence. Thus the welfare of the potential child cannot be the basis of claims regarding benefitting or harming the child; therefore, the considerations that are used to restrict access to reproductive technology and service are unfounded. This becomes even more apparent when the denial of access to reproductive technology services means that no child ever exists. As the child never exists, it is logically impossible for them to be ‘benefitted’ or ‘harmed’, but this also applies in reverse, that is, a child who does come to exist cannot be benefitted or harmed before their existence either. This is what Heyd calls ‘full symmetry’ between benefitting and harming in both cases of not bringing and of bringing a child into existence.

⁵ HFE Act, section 13(5).

⁶ David Heyd, *Genethics: Moral Issues in the Creation of People* (University of California Press, 1st ed., 1992) 36.

Consequently, the law relating to reproductive technology and services ought to be reformed by removing the ‘welfare of the potential child’ test from the HFE Act, as it is based upon a concept of harm in which harm is impossible. Using the example of genetic technology⁷ available to parents (such a genetic testing, screening, selection etc.⁸), the implications of this test for what is permissible will be explored, and the scope of parents’ decisions will be explicated. These implications do not however suggest that parents should have unrestricted access to these services (because resource constraints and distributive justice are also involved), but they do mean that parents cannot be denied access to genetic modification services on the basis of harming a child who will never exist and who cannot be harmed if they do exist. Only actual persons can be affected by the actions of the parents; for example, through increasing burdens on health care services by allowing parents to have disabled offspring. Furthermore, the welfare of the child once it actually exists is covered by the ‘welfare of the child test’ found in section 1 of the Children Act 1989, thus providing a means for dealing with parents if they are abusive to their offspring. Society could therefore decide that people who want to be parents have to conform to a certain standard of conduct and comport themselves in a certain manner, but this would have to apply to all parents (whether using reproductive technology or conceiving naturally) because the determination and judgement is located in the parents themselves⁹. Currently, those seeking medical assistance to have children are subject, in effect, to a competency type test which is not applied to the rest of the population. The arguments that will now be presented will support the final recommendations of this article, namely that the welfare of the potential child test in the HFE Act must be removed and that the determination of access to reproductive

⁷ Genetic technology here encompasses all the actions parents might take to ensure that their future offspring either possesses or does not possess a particular genetic trait: see ‘New Down’s syndrome blood test more reliable, say researchers’ *The Guardian*, 7 June 2013; ‘Lesbian couple have deaf baby by choice’ *The Guardian*, 8 April 2002; ‘IVF baby born using revolutionary genetic-screening process’ *The Guardian*, 7 July 2013; ‘Families hope “Frankenstein science” lobby will not stop gene cure for mitochondrial disease’ *The Guardian*, 5 February 2014.

⁸ For more work on the issues of genetic technology, see Allen. Buchanan et al, *From Chance to Choice* (Cambridge University Press, 2001); Jonathan Glover, *Choosing Children: Genes, Disability and Design* (Oxford University Press, 2008); John Harris, *Enhancing Evolution: The Ethical Case for Making People Better* (Princeton University Press, 2007); Loane Skene (ed.), *The Sorting Society: The Ethics of Genetic Screening and Therapy* (Cambridge University Press, 2008).

⁹ As far as I am aware, no one has argued that all parents should prove their competency to raise children. There are a number of programs, such as the Triple-P program (Positive Parenting Program), which are designed to give parents strategies for raising children, although these are obviously voluntary.

services should be based upon the effect on actual persons or a standard applicable to all.

The welfare of the (potential) child test

Under section 13(5) of the HFE Act, the welfare of the child test operates as follows:

A woman shall not be provided with services . . . unless account has been taken of the welfare of the child who may be born as a result of the treatment (including the need of that child for supportive parenting), and of any other child who may be affected by the birth.

The HFE Authority's 8th Code of Practice guidance notes on the welfare of the child state that the 'factors to be taken into account during the assessment process' of the parents include the risk of 'serious physical or psychological harm or neglect'¹⁰ and 'where the medical history indicates that any child who may be born is likely to suffer from a serious medical condition'. Subsections 13(9) and 13(10) of the Act, inserted by the amending 2008 Act, state that 'embryos known to have' a serious disability, illness or 'other serious medical condition'¹¹ 'must not be preferred to those [embryos] that are not known to have such an abnormality'.¹² Together, these legislative provisions and code of practice notes make two things clear (1) that the conduct of parents is subject to scrutiny and (2) that parents cannot select for an embryo that has a disability but can reject an embryo on that basis. As both sets of considerations operate under the welfare of the (potential) child test, it is clear that the reasons and justifications for these considerations are rooted in assumptions that they are harmful to the child's welfare.

Thus a positive determination that the potential child will be subject to abuse or neglect, or will have a serious disability, will prevent the parents from accessing reproductive services, frustrating in its entirety their interest and desire to have children. This places parents in the invidious position of not being able to have children without medical help but also of being judged both on their characteristics and their choice of foetus. Moreover, if any of these judgements go against them, then

¹⁰ Human Fertilisation and Embryology Authority, Code of Practice, 8th ed., 2009, section 8.10, Available at: <http://www.hfea.gov.uk/3401.html> (accessed 21 July 2014).

¹¹ Code of Practice, section 8.10(b)(iii).

¹² Human Fertilisation and Embryology Act 2008, section 14.

the child they would supposedly be harming would never even exist. Persons who do not exist cannot be harmed because their non-existence precludes anything affecting them in any sense. We will come to the impossibility of harming potential persons in due course. For now, we can note the effects of the current legal regime. Firstly, by requiring those who seek reproductive services to prove their ability to be ‘good’ parents, those who are already at a severe disadvantage are further penalized. The institutional burden imposed on them, which is not imposed on other members of society, could be construed as an unacceptable interference with the parent’s reproductive autonomy. Some have argued that a right to reproductive autonomy can be found under Articles 8 and 12 of the European Convention of Human Rights, and interference with this right would need to be strongly justified. The discussion of reproductive autonomy, and whether there is a right to it, is beyond the scope of this article.¹³ However, it should be noted that the arguments presented here mean that the requirement of proving parental capability would not make sense on the basis of the potential child’s welfare. Requiring proof of parenting competence could be justified if applied to all potential parents, but it would have to be a universal requirement based upon the interests of actual members of society. Thus the justification for the burden of proving parental capability would be focused solely on the parents and would not be related to the supposed characteristics of the potential child.

More importantly, though, the demand that disabled embryos are not to be preferred and the fact that disability counts against the potential child’s welfare implies that there is an obligation not to have disabled children. This obligation can be inferred because the logical corollary of the statement that disabled fetuses ‘are not be preferred’ is that non-disabled fetuses are to be preferred.¹⁴ This restriction on having disabled offspring removes the choice of parents and mandates what they can and cannot do. Furthermore, if genetic modification could improve the potential child’s ‘welfare’, then it may be required of parents to use genetic technology because the basis of the constraints are the potential child’s interests which (as shown by the

¹³ For the legal consideration of a right to reproduce, see *Dickson v. UK* [2008] 46 EHRR 41; *Evans v. UK* [2008] 46 EHRR 34. For further discussion, see Marleen Eijkholt, ‘The Right to Found a Family as a Stillborn Right to Procreate?’, 18 *Medical Law Review* (2010) 127; Maja Kirilova Eriksson, *The Right to Marry and to Found a Family: A World-Wide Human Right* (Coronet Books, Inc, 1990); Emily Jackson, *Regulating Reproduction* (Hart, 2001); Sheila McLean, ‘The Right to Reproduce’, in Thomas Campbell and others, eds., *Human Rights: From Rhetoric to Reality* (Basil Blackwell, 1986) 103.

¹⁴ Human Fertilisation and Embryology Act 2008, section 14.

restrictions imposed on parents by the HFE Act now) are strong enough to force parents to refrain from acting (and thus logically can require them to act) in certain ways.

As the possibility of requiring parents to use genetic modification of the embryo would be justified on the basis that disability is against the potential child's welfare then, if this justification is incoherent, the preference for non-disabled offspring and the constraint imposed on the parents must be removed. Parents could therefore not be prevented from having disabled offspring if they so wished, at least on the basis that it is against the child's interests. The welfare of the potential child then is the source of the problems that are thrown up by applying the welfare test to those seeking access to reproductive technology; namely that the potential child welfare acts as justification for restricting who can have children and under what conditions, when in fact the child's welfare cannot be affected by any decision made. It is to this person-affecting dilemma to which we now turn.

Welfare of the child: What child?

The fundamental problem with the welfare of the (potential) child test is that welfare by definition is for someone, but in this case the person concerned does not exist. The welfare of the child test, contained in the HFE Act section 13(5), explicitly includes the concept of harm to potential persons, thus identifying the potential child as the subject affected. This is, however, a logical impossibility because the necessary condition for welfare to be considered, the existence of a person, is not met. Moreover, if the potential person's welfare is considered to be at risk, then they will never come to exist so they will never benefit from this welfare decision. As welfare is 'person affecting', it is predicated upon the existence of someone for whom welfare is for, but this person must therefore be able to value or disvalue things which affect their welfare; that is, they must be a 'valuer'. As value cannot exist without valuers (in this case, welfare cannot exist without being for someone), then 'valuers have the unique status as the condition of there being any value'.¹⁵ In the cases under discussion (prospective parents seeking fertility treatment), the only valuers for whom the decision can be of value are the parents. The potential child therefore cannot have

¹⁵ Heyd (n6) 117.

a 'welfare concern' that can be affected when it is the decision which will determine whether it exists or not.

This seems clear and highlights the fact that deciding against the existence of potential person cannot count as beneficial to them. Obviously it could be beneficial to the parents as existing valuers, because they would not have the burden of raising a (disabled or non-disabled) child, and it could be beneficial for other actual members of society, for example, due to the consumption or distribution of health care resources. But it is painfully obvious that it cannot be beneficial to the potential person when they will never exist. For example, genetically modifying an embryo so that it no longer has a genetic disability would not benefit it, because the decision to genetically modify would constitute the potential person; that is, it would be a decision prior to and determinative of their existence. There is no basis for the claim that offspring are harmed by genetic modification when it is that genetic modification which causes them to exist.

Furthermore, 'equally and for the same reasons, we cannot hold the child to be an object of maleficence'.¹⁶ Just as a valuer needs to exist before they can benefit from things so they have to exist before they can be harmed by things. For example, a potential disabled child cannot be harmed by being brought into existence. Thus 'potential people have equal standing regarding our beneficence and our nonmaleficence', which is to say they have no standing.¹⁷ Therefore there is no distinction between obligations to benefit or to not harm a potential person; neither type of obligation is engaged. However, this also means that creating new valuers is itself of no value independent of those valuers who already exist, there thus can never be any presumption to implant an (or any) embryo on the grounds of benefitting the potential person created by this process. This also means that there can never be a presumption against implanting an (or any) embryo justified by avoiding harm to the potential person. It is only on the basis of the interests of actual persons, the parents, that decisions regarding offspring can be made.

Once the potential person does exist then they can be affected by the circumstances in which they exist. The necessary precondition for value (harm and benefit) to attach to someone has been met (by their existence) and from there onwards they possess interests because they begin to form a 'biographical life'. A

¹⁶ Ibid 109.

¹⁷ Ibid 115.

biographical life consists of the actions, desires, projects and decisions that make up a life; that is, the components that give a life consistency.¹⁸ It should thus be self-evident that an existing child has interests that can be accounted for and protected, and these post-existence effects can form the basis of welfare decisions. At this point, the Children Act 1989 section 1 would be applicable and would protect the actually existing child, further eroding the purpose and need for the welfare test in the HFE Act. The case of disability is a good example of the argument outlined above. If an embryo has the genetic characteristics of being disabled, the HFE Act considers the disability to count as a reason for rejecting that embryo and prohibits selecting that embryo when a non-disabled embryo is available.¹⁹ This is so even though the offspring would not actually come into existence because the embryo has been rejected and thus no subject would ever exist to benefit or suffer harm from the decision taken. However, this also applies to a decision to allow the disabled embryo to develop to term, because the potential person must exist before the 'person-affecting' nature of value can be present. In such cases (rejection, acceptance or modification of the disabled embryo), the potential person cannot be affected by the decision taken, but the parents can. Thus, it is their interests which are relevant when deciding which embryo to implant and it is up to them to determine whether having offspring with a disability is against or in their interests. Once a child exists (whether disabled or not) then the precondition for value to be for them has been met and they will possess interests that can be benefited or harmed. The welfare of the child test under the Children Act will now apply and judgements regarding the child's welfare can now be made, thus forming the basis for decisions and actions about the child's upbringing and care.

As noted throughout this article, the basis of removing the potential person (the child who may result) from consideration is the necessity of their existence before they can be harmed or benefited. This applies equally to negative constraints (such as not selecting certain embryos) and positive requirements (such as genetically modifying an embryo). Consequently, it is impossible to maintain the HFE concept of the welfare of the offspring who results from the treatment because potential people have no welfare to speak of. The welfare of the (potential) child test should therefore

¹⁸ Statement by John Harris (personal communication, 12 April 2013).

¹⁹ Human Fertilisation and Embryology Act 1990 as amended, Human Fertilisation and Embryology Authority Code of Practice, 8th ed., 2009.

be abolished and the consideration of the potential child removed as a factor for consideration when deciding on distributing reproductive technology and services. Any restrictions or constraints imposed upon parents would then have to be based upon the impact on actual persons which would include the parents seeking reproductive aid and the other actual members of their sociopolitical unit (more on this later). The general form of the test found in the Children Act, with its focus on existing offspring, would, of course, retain its applicability to offspring once they exist.

We can thus see how this change would eliminate the restriction on parents when it comes to the selection of different embryos for implantation. Next, we will consider an alternative to the person-affecting approach, that is, impersonalism. If there is an alternative to the claim that welfare must be person affecting, then this may undermine the claim that welfare must be for someone. This alternative, however, can only be based on a non-person-affecting account, namely, impersonalism.

Impersonalism: The alternative to the person-affecting approach

As has been made clear throughout this article, the welfare of the (potential) child is logically unsustainable if welfare is attached to a subject. Thus, if welfare can be framed in such a way that it does not attach to a subject, this would offer another method for regulation to incorporate the welfare test. To be clear, impersonalism here is not ‘a common standard’ that is good for everyone; as this would simply be claiming that there is some x which is good for all subjects who exist.²⁰ It would thus still be tied to the actual existence of people but would simply be a claim that x is good for all of them. This obviously would not be an alternative to, or an answer to, the argument set out above because it would still require that a person exist for them to benefit from the common standard. Here, impersonalism is used to mean the existence of value as a feature of the world; that is, value is independent of persons and attaches ‘to the world’.²¹ This stance would thus identify what is good and claim that the more of this good, the better because it is as a feature of the world that it has value, not as it relates to persons (if they exist).

So what might an impersonalist approach look like? Julian Savulescu argues for an approach in which the offspring ‘who is expected to have the best life, or at

²⁰ Heyd (n6) 82.

²¹ Ibid 80.

least as a good as life as the others' should be chosen.²² We would therefore not select a disabled foetus because it would not have the best expected life. However, as Rebecca Bennett puts it, 'This individual is born in this impaired state or not at all', and thus the choice is between the existences of different potential people and on a person-affecting account this raises the same problems outlined above.²³ Consequently, we could not say it is better for those who will not exist, as this would be to take a person-affecting perspective. Of course, on a person-affecting view, we could say that it is better for those of us who actually exist that this particular potential person does not exist. For Savulescu's comparison to work, it would have to be based upon impersonal criteria, particularly as the existence of each potential person is mutually exclusive of the others. In order to avoid relying on the effect on any people, what we would have to say is that the world is a better place with those particular potential persons in it, rather than these other potential persons, because those we allow to exist will contribute more good (in this case the best lives) to the state of the world. In other words, the world is better off than with alternative persons in it because they add to the total good in the world. Alternatively, it could be that impersonalism is based upon being the right kind of person (i.e. being human consists of suffering and limitation).²⁴

The essential problem with impersonalism is that it 'sacrifices the utility of individuals to the promotion of the impersonal value of the overall good'.²⁵ In other words, whatever the good is, people are merely a means to produce the good, and thus individuals are always expendable. This means that people are not relevant beyond the production of the good. For example, if health was the primary good (as suggested in Savulescu's scenario), then all parents could be forced to undergo whatever procedure maximized health, as long as the cost in the health of the parents was outweighed by the gain in the offspring. If impersonalism is based upon creating the right kind of people then this balancing act would be unnecessary; it would only be when the wrong kind of offspring were to be produced that parents would have to take measures to prevent it whatever those measures are. Now it is obvious that if

²² Julian Savulescu, 'Procreative Beneficence: Why We Should Select the Best Children' (2001) *Bioethics* 413, 413.

²³ Rebecca Bennett, 'The Fallacy of the Principle of Procreative Beneficence' (2009) *Bioethics* 263, 267.

²⁴ Leon Kass, 'Ageless Bodies, Happy Souls: Biotechnology and the Pursuit of Perfection' (2003) *The New Atlantis* 1, 9.

²⁵ Heyd (n6) 57.

impersonalism is the correct approach, then the subjugation of individuals to the production of the good or to the creation of the right kind of person is not only permissible but an obligation. The restrictions imposed upon individuals could thus be justified.

But impersonalism is not a sound or coherent theory because the identification of the good or the right kind of people unavoidably presupposes the existence of valuers. The good that forms the basis of impersonalism (whatever it is) is 'by necessity of actual' persons.²⁶ Without actual persons, that is valuers, it is impossible to differentiate between different conceptions of the good because a world without persons is valueless. Otherwise anything could be good, even the murder of humans could be justified on the basis that it is more natural for a predator species to behave in such a manner and thus promoting the good of the natural world. Thus, the existence of persons who can identify the good is a prerequisite for impersonalism to work, but this is the very element that impersonalism claims to do without. More to the point, impersonalism itself gives all individuals a reason to reject it because it could impose upon them a good which they do not find to be good. Impersonalism can therefore only work at one level of discourse, namely the identification of objective goods for a particular reference group. It cannot be used to justify the existence of members of that reference group because the identification of value without people is impossible. Furthermore, the identification of a good to which persons are subservient is not possible, because good has to relate to the person or persons affected.

Thus we cannot resort to impersonalism to solve the conundrum of the potential child possessing welfare interests, because impersonalism reduces persons to mere producers of a good and thus justifies any restriction or interference which promotes this good. It is also conceptually incoherent as the good cannot be identified without presupposing the existence of the reference group (valuers). Once the reference group exists, then the conception of the good becomes person affecting, as only by tying good to valuers does it make sense to differentiate between potential goods. The arguments pertaining to the illogical and impossible conception of welfare for a potential person remain valid and cannot be replaced with impersonalism. We therefore have to accept that the welfare of the potential child test is unsustainable and

²⁶ Ibid 63.

that the law is wrong to have such a test because it is factually impossible for the test to have any effect on the person it supposedly protects. The only option left is to remove the test completely. This does not however mean that parents have or should have unfettered decision-making powers over reproduction, but it does mean that such restrictions must be based upon the interests of actual persons; in this case, the parents and the other members of their socio-political unit. It is to this consideration that we now turn.

The decisions of prospective parents

Having dispensed with the welfare of the potential child test, we need to decide if and on what basis the reproductive decisions of prospective parents could be constrained. As the preceding section pointed out, without the welfare test and the requirement that the decisions take a person-affecting form, reproductive ‘genesis choices . . . should be guided exclusively by reference to the interests, welfare, ideals, rights and duties of those making the choice’.²⁷ Those who are making the choices are the ‘generators’ of the (pro)-creative effort and, as it is their interests that are of concern, Heyd calls this theory of their predominance ‘generocentrism’.²⁸ Generocentric theory will form the foundation of the new reformed approach to reproduction, specifically in the case of parents utilizing reproductive technology and services, where the decisions of the parents are contingent upon the acceptance of those providing the service for the decisions to take affect. Artificial reproduction therefore provides a case in point where the interests of other actual persons and the interests of parents intersect.

From the arguments set out throughout this article culminating in generocentric theory, there can be no restrictions imposed upon parents from the perspective of the potential child. Parents should therefore have greater control over the reproductive process unconstrained by the supposed impact on the potential child. What is of interest here is what possible constraints there might be for the parents in the absence of the welfare of a potential child. As pointed out earlier, any such constraints would have to be based upon either the interests of other people or upon a requirement for the parents to meet a standard of conduct. Given the emphasis on the person-affecting nature of actions, an approach grounded upon the interests of other people is the only justification for restricting the actions of parents, but this could of

²⁷ Ibid 96.

²⁸ Ibid 96.

course take the form of a standard of conduct. However, we will avoid becoming encumbered by discussions between the relative merits of the different forms the restrictions may take. We shall instead focus on the effects that would justify imposing the constraints.

Some examples are the increase of health care costs; ensuring genetic diversity for the gene pool; preventing hyperaggression; improving immune systems, muscle strength, liver function, and brain functioning and so on. The justification for restraints could also relate to the kind of society that we are trying to build, whether it is strongly democratic and pluralistic with a strong sense of individualism, whether strict conformity to a particular ideal is required or whether reproduction is considered a collective effort or interest. Given that the jurisdiction under discussion is the United Kingdom, this article takes it as given that the constraints have to be compatible with a democratic pluralist human rights respecting society. In this case, the basis of restrictions has to be proven by those seeking to restrict the actions of the parents. This would be even stricter if a formal right to reproduction or reproductive autonomy were to be created or recognized, because the restriction would also have to be necessary and proportionate. Unfortunately, the decision as to whether there should be such a right is beyond the scope of this article.

Thus, it must be noted that this person-affecting approach would entail a strong form of reproductive autonomy for parents generally in the area of reproduction as the impact upon others would have to be proven. For example, if someone claims that it would be harmful to have a disabled child then they must show to what interests it is harmful; such a claim could not be based upon simply disapproving of the choice. It is therefore plausible that only a significant interest could restrict parental decisions. Such a significant interest might be the interest of actual persons to have access to health care. Thus, parents who choose to have disabled offspring might be constrained on the basis that the additional resource burden on the health care system and the subsequent effect on extant people's interests make the choice to have disabled offspring an unreasonable expectation for the use of public resources. It could not, however, be based upon the harm done to the parents themselves, because it is up to them whether the burden of raising disabled offspring is outweighed by the benefits of having a child, but it could easily be based upon the effect on others. Such a justification would encompass concerns of fair distribution of resources, safeguarding and balancing conflicting interests (and/or

rights) among a population, population growth policy and societal aims, changes or objectives. One scenario may be that the interests of actual persons make it preferable to ensure a universally accessible standard of reproductive technology rather than a private system that only some could afford.

Other modifications however may be problematic; skin colour, for example, may be something that remains outside of parental remit, not because it may create a burden or thwart the interests of others but because allowing such a choice in a society which unjustifiably discriminates against certain skin colours may cause reproduction (and genetic modification) to be used to reinforce racism and other prejudices. Such a consequence could be opposed because diversity is valued by society as leading to a richer and deeper culture, because it is a superficial difference or because the promotion of a united human governmental system or human rights culture are considered important for improving the lives of actual persons. In this case, the justification could be based upon the harm done to actual persons in the sense that they would live in a society that is worse for them because it violates the basic principles or objective of society or because it worsens the situation of actual persons. Deciding on what may and may not be permitted is up to the sociopolitical unit where artificial reproduction takes place; it is beyond the scope of this article to consider the political methods which should be used. However, in no case could it be based upon the potential person themselves, as it is impossible for them to have interests and because they are irrelevant themselves except for their value to extant persons.

Finally, imposing constraints on parents may be justified because we (i.e. the rest of society) do not want them to achieve or satisfy their interests. For example, parents who want to inflict suffering and harm on a person may be prohibited from utilizing reproductive technology and services because we do not want to help people satisfy their desires to act in cruel and abusive ways by creating a child. In such cases, we are not saying that the child's interests would be harmed by being created for such a purpose, and once they exist, their interests would justify removing the child from the parents, although this may be a difficult and protracted process. Rather, we are saying that the interest the parents seek to fulfil is prohibited because society seeks to minimize the prevalence of cruel behaviour among actual persons; thus they cannot be permitted to achieve their goal. This would be similar to not allowing people to vandalize public or abandoned property, to torture willing 'victims' or to take class A

drugs even if no one is directly harmed by such conduct. This is because some conduct is deemed unacceptable by society and those who would use reproductive technology and services to fulfil these prohibited objectives could be prevented from doing so. Obviously these prohibitions purport to represent more than just likes and dislikes of society; how we might determine which rules to adopt would have to be underpinned by a full a normative theory.²⁹

Regardless, within the United Kingdom and other democratic pluralist human rights respecting nations, the interests and decisions of the parents would have priority. This approach would therefore create a legal presumption or right in favour of permitting prospective parents to reproduce the offspring they wish, and those who wish to restrict the actions of parents would have to make their case and provide evidence of their claims. Thus the person-affecting approach does not preclude restricting the actions of prospective parents on the basis of the interests of other persons or on the basis that prospective parents should act in certain ways. This would be based on competing interests between societies, existing members and the parents. This provides the justification for judicial oversight of reproductive technology whilst protecting parents from being denied access to reproductive services on the basis of the welfare of the potential child.

Conclusion

The fundamental argument of this article is that potential persons have no standing (moral, ethical or legal) in relation to those who actually exist. Consider the following, imagine that every person on the planet joins the Voluntary Human Extinction Movement³⁰ and through a program of permanent and irreversible sterilization the human species ensures that it cannot reproduce by any means. In such a case, all potential persons would cease to be potential at all, because none of them will ever exist. We could not say that we owe a duty to these non-existent beings of a no longer possible future; and if we do not owe an obligation to them to ensure they exist, then we cannot owe them any duty to have a particular life. This also has implications for environmental strategies, as it would mean that we cannot owe them a duty to provide any particular environment or to provide as certain level of

²⁹ For those who wish to consider this point further, contractarian theory provides a good example of developing rules based upon a normative system of rational agreement.

³⁰ Available at: <http://www.vhemt.org/> (accessed 29 August 2014).

resources. A potential person cannot benefit from existing or not and neither can a potential person be harmed nor are they owed any obligations.

Consequently, when it comes to reproduction, the interests of prospective parents are unconstrained by the effect of their actions on their potential offspring because actions which cause a person to exist cannot affect them. Parents are thus free to choose to have disabled, genetically modified or 'normal' offspring as they wish. The only source of restrictions on parental conduct comes from the effect of their actions on other actual persons. This approach is explicitly person affecting which has as a precondition for it to apply that a person actually exists. It is this that underpins the argument that potential persons have no standing compared to the interests of actual persons, be they parents or members of society.

The importance of this change in outlook is prompted by the inclusion of the welfare of the (potential) child test in decisions about access to reproductive services by infertile couples. This is objectionable because those already experiencing difficulties in having children are further penalized by having to prove their capacity to be parents; and all this is because of a potential person who cannot be affected by their actions. The point at which the potential offspring becomes actual (when the child can be affected) is covered by the Children Act 1989. Thus, parents have their choices, decisions and actions restricted by the interests of a being that does not exist. When we consider the impossibility of harming potential persons, we can see that this is an unjustified restriction imposed upon parents which consequently should be removed. Only by taking an impersonalist non-person-affecting approach could we avoid this difficulty.

Impersonalism has however been shown to be incoherent and flawed because it is abhorrent to individuals by making them into producers of value (reducing their own interests into irrelevance) and because it presupposes the existence of valuers in order to determine and identify the value(s) that are to be produced. Note once again that this is not to say that impersonal (objective) considerations are irrelevant when considering benefits across an identified reference group but that the very foundation of benefit, harm and value cannot be disconnected from the existence of evaluators. Thus impersonalism needs persons to exist before it can have any meaningful content, but it then discounts the interests those persons have once they do exist. The alternative to the person-affecting approach is therefore untenable, and this brings us

full circle back to the determination of the welfare of the potential child test as unacceptable and leaves us with only the interests of actual persons to consider.

Thus parental decisions would be unrestricted by the concept of harming their future offspring as those actions are the ones that bring the offspring into existence. We should note however that it has recently been shown that there is a general and widespread uncritical acceptance of the requirement and that the regulatory regime operates in a 'light touch' manner.³¹ This suggests that because there is such widespread belief, enforcement can be light touch. Without this rule, as proposed by this work, then people may still act on the basis of concern for the welfare of future persons. This suggests that the existence of the rule may be unnecessary and could be interpreted as supporting my position. On the other hand, it could be argued that if such belief is widespread then it should be reflected in the legislation. Either way it calls for further empirical analysis of service users and providers experiences in artificial reproduction.

Parents would thus be able to utilize reproductive technology (including genetic modification) with only legitimate public interest limiting their actions. Of course, medical providers will have obligations to the parents, and parental actions will be restricted by the technological capabilities of the time. Moreover, depending on how services are provided, for example, by a public or private system, parental decisions will be limited by different constraints. What is made clear is that parental interests have primacy and can only be overridden by the interests of other existing citizens and not by the interests of the potential offspring. This could lead to a legal presumption that parental decisions to have disabled offspring or genetically modify their offspring are in their interests and that if others want to prevent such modification going ahead, then the burden is on them to prove that there is a sufficient public interest or harm being caused to actual persons.

As we have seen, this may take a number of forms but can be based on the harms caused directly to other persons, the fairness and distribution of resources or on the ideals a society holds. For example, we want to prevent those who sadistically want to torture people from achieving their aims, or we want to ensure that every member of society has access to services on an equal basis and so some limitations to

³¹ See Ellie Lee, Jan Macvarish and Sally Sheldon, 'Assessing Child Welfare Under the Human Fertilisation and Embryology Act 2008: A Case Study in Medicalisation?' (2014) 36(4) *Sociology of Health and Illness* 500.

the number of times reproductive services are accessed is imposed. Another example may be prohibiting the selection of skin colour because we do not want to exist in a more prejudiced society than we do now. All of these examples, whether to do with the harm caused to a specific individual, the fairness of health care access across the entire population, population control policy or because of the particular ideals a society holds, are based upon the impact and effect of actions on existing actual persons. Restrictions and constraints imposed on this basis are therefore permissible and justifiable. However, depending on what form the parents' interests take, for example human rights, will affect the strength needed for justification and the degree to which interference can extend.

The reform recommended here, that is, the removal of the welfare of the (potential) child test in relation to artificial reproductive services, would clearly make the interests of the prospective parents paramount. As the parents are the ones whose interests would be fulfilled by reproduction and because it is such a major component of their lives, it is fair that decision is accepted unless it can be shown that their decision is harmful to an actual person or because it violates the standards of conduct held in the socio-political unit. A decision as to whether parental discretion should be embodied in a human or legal right is a discussion for another time. It is clear, however, that the injustice and unfairness imposed by the welfare of the (potential) child test is untenable and should be excised as part of the process of artificial reproduction. If the actions of prospective parents do not harm anyone or violate behavioural codes then they should be free to reproduce as they wish.

Subsection 4B: Impersonalism and the creation of individuals

Since this chapter has been published I cannot change the substance of the paper. I do, however, now think that I need to present a more in-depth analysis of the problems inherent in defining ‘harm’ and especially analyse a number of further issues relating to the impersonal / person-affecting harm distinction. In this subsection I will outline the points that I would now make concerning the welfare of the child test and its reliance on the notion of impersonal harm. This section aims to show two things. That impersonal and person-affecting harm are mutually incompatible and, while each of these views is axiomatic, person-affecting harm is empirically demonstrable while impersonal harm is a cognitivist claim about the world which can be rejected. I begin by setting out how attempts to avoid the conclusions of person-affecting harm must utilise a cognitivist notion of impersonal harm that can be applied to individuals regardless of their subjective experience. This leads us to those factors that have been proposed by various scholars to function as this objective measure of impersonal harm but when we analyse these proposals it becomes clear that these factors are subjective selections made by the individual theorists. I therefore conclude that impersonal harm is used to justify the particular prejudices of each author rather than identifying something that is objectively applicable to any and all agents. However, I further conclude that even with these problems individuals may still adopt an impersonal viewpoint about harm but this cannot justify imposing that view on those who reject impersonal harm. In contrast person-affecting harm can be shown to affect agents and so is an unavoidable factor in interaction but this requires existence and so is only applicable to extant beings. Consequently, we should not use impersonal harm to justify population wide constraint but only person-affecting harm which cannot apply to potential persons. In the case of reproduction, this means that the welfare of the potential child cannot be a factor except where individual agents adopt the perspective of impersonal harm.

The non-identity problem arises when one tries to maintain a person-affecting account of – in this case – harm in relation to the creation of people (henceforth genesis cases). Non-identity is a problem for person-affecting harm in genesis cases because our decisions *affect who* will exist, but do not *affect* those who *do* exist. When

preventing harm from occurring prevents the existence of a person then no one is made better off by that action. Thus the non-identity problem points out that:

Identity-affecting decisions are those which affect not what life will be like for a fixed future population or person, but instead affect which persons (out of a set of possible future persons) come to exist in the future.¹

This is because:

Which particular future people will exist is highly dependent upon the conditions under which we and our descendants procreate, with the slightest difference in the conditions of conception being sufficient, in a particular case, to insure [sic] the creation of a different future person.²

In genesis cases, because our actions change the identity of those who will exist, it is impossible for us to improve or worsen the situation of those who exist as a result of our actions – our choice is between two distinct possible lives, not between two possible states of one person.

Non-identity is particularly a problem for the current legal system's *welfare of the child test* under the Human Fertilisation and Embryology Authority because welfare considerations can only apply to existing children. The Human Fertilisation and Embryology Authority's (HFEA) 8th Code of Practice explicates the notion of welfare by specifying the content of the test as follows:

8.3 The centre should assess each patient and their partner (if they have one) before providing any treatment, and should use this assessment to decide whether there is a risk of significant harm or neglect to any child.³

¹ Stephen Wilkinson, *Choosing Tomorrow's Children* (Oxford University Press 2010) 69.

² Gregory Kavka, 'The Paradox of Future Individuals' (1982) 11 *Philosophy and Public Affairs* 93, 93.

³ Human Fertilisation and Embryology Authority, '8th Code of Practice' (2012) s.8.3.

Furthermore, the Human Fertilisation and Embryology Act 2008 (the 2008 Act) inserted s.13(9) into the Human Fertilisation and Embryology Act 1990 (the 1990 Act) directs that:

[E]mbryos that are known to have a gene, chromosome or mitochondrion abnormality involving a significant risk that a person with the abnormality will have or develop—

(a) a serious physical or mental disability,

(b) a serious illness, or

(c) any other serious medical condition,

must not be preferred to those that are not known to have such an abnormality.⁴

Clearly, the current regulatory regime operates on the basis of a person-affecting notion of harm. But a welfare test based upon the person-affecting notion of harm is unjustified because only once potential people exist do ‘they satisfy the logical condition of a “person-affecting” view’.⁵ Until they do satisfy these conditions ‘potential people have equal standing regarding our beneficence and our nonmaleficence’⁶ – which is to say they have no standing – thus we cannot hold potential persons ‘to be an object of maleficence’.⁷ Here, I argue that only an impersonal account of harm can apply to genesis cases but this applicability comes at the expense of justifying constraint on a population of heterogeneous agents.

Impersonalism, Objective Values, Hidden Subjectivity and Error Theory

An impersonal account of harm assigns value to the suffering present in different lives permitting a ranking of all lives and would make sense of the claim that lives have negative value or fall below a minimum amount of positive value. Additionally, impersonal harm, by providing a ranking of lives, would simultaneously justify positive selections of better lives. Consider Julian Savulescu’s approach in which the offspring ‘*who is expected to have the best life, or at least as good a life as the others*’

⁴ Human Fertilisation and Embryology Act 2008 s.14(4) (my emphasis).

⁵ David Heyd, *Genethics Moral Issues in the Creation of People* (University of California Press 1992) 86.

⁶ *Ibid* 115.

⁷ *Ibid* 109.

should be chosen.⁸ An impersonal metric would identify those lives which are better and those that are worse based upon the characteristics of those lives. Of course, this is not to say that it is a better life *for those who will exist* as they do not have an alternative life – because ‘[t]his individual is born in this impaired state or not at all’ thus each individual has the best life possible *for them* – rather the claim is that it would be a better life *per se* because those lives have more of whatever objective property it is that makes lives valuable.⁹

One of the challenges for impersonalism is identifying the criterion to be used to assess the quality of individual lives independent of any effect upon the individuals whose lives they are. An analysis of Savulescu’s work demonstrates the difficulty in specifying a metric for impersonal harm. He suggests, for example, that the ‘relevant property of ‘asthma’ is that it is a state which reduces the well-being a person experiences’.¹⁰ He goes on to say, ‘[b]y ‘best life’, I will understand the life with the most well-being’.¹¹ This implies that ‘well-being’ is the concept that justifies the inter-person comparisons but this is confused by the phrase *well-being a person experiences*, which suggests that a person-affecting notion applies. In later works Savulescu seems to add a further claim that ‘certain traits will increase the choices and opportunities of our children because they are opportunity-enhancing’.¹² In another paper he (along with Guy Kahane) seems to include two additional accounts as well:

According to the wide person-affecting version, our reason to select the child with better prospects is that that child will benefit more than the other would by being caused to exist. According to the impersonal version, our reason is that selecting the most advantaged child would make the outcome better, even if it is not better for the child created.¹³

⁸ Julian Savulescu, ‘Procreative Beneficence: why we should select the best children’ (2001) 15 *Bioethics* 413, 413 (original italics).

⁹ Rebecca Bennett, ‘The Fallacy of the Principle of Procreative Beneficence’ (2009) 23 *Bioethics* 263, 267.

¹⁰ Savulescu ‘Procreative Beneficence: why we should select the best children’ (n8) 417.

¹¹ *Ibid* 419.

¹² Julian Savulescu, ‘Procreative Beneficence: reason to not have disabled children’ in Loane Skene and Janna Thompson (eds) *The Sorting Society* (Cambridge University Press 2008) 51, 65.

¹³ Julian Savulescu and Guy Kahane, ‘The Moral Obligation to Create Children with the Best Chance of the Best Life’ (2009) 23 *Bioethics* 274, 277.

It is unclear which of these claims Savulescu would favour but the well-being, opportunity-enhancing and wide person-affecting notions all fall victim to Rebecca Bennett's counter-argument (to which it seems Savulescu has never responded) that the child who *actually* comes to exist will have the best life it can possibly have whilst the other possible children never exist. As Bennett puts it:

[Whilst] we wrong an individual by knowingly bringing him to birth into a life overwhelmed with suffering [the life worth living threshold], we do not wrong an individual by bringing him to birth into a life which is likely to be of positive overall value.¹⁴

As the child who actually comes to exist can only have *that life*, it is not harmful or wrong to bring that child into existence as long as that life has an overall positive value – that is it meets the life worth living threshold. Only on an impersonal account does it seem that inter-life comparisons are possible and Savulescu has failed to provide a means for interpersonal comparisons – particularly as all of his proposed criteria are inherently subjective judgements about what makes life go well.

This problem also plagues the concept of a *life worth living* for this also requires an impersonal metric.

A life not worth living is not just worse than most people's lives or a life with substantial burdens; it is a life that, from the perspective of the person whose life it is, is so burdensome and/or without compensating benefits as to make death preferable.¹⁵

Obviously, the *perspective of the person whose life it is* must not to be taken literally as this would require that an agent exists, instead it must be taken to be hypothetical. *If the person whose life it would be existed they would prefer never to have been born.* The life worth living threshold has two components. First 'it suggests that life can be a net burden to a child if the disadvantage is severe enough' and second 'it maintains

¹⁴ Bennett (n9) 267.

¹⁵ Allen Buchanan, Dan Brock, Norman Daniels and David Wikler, *From Chance to Choice* (Cambridge University Press 2000) 224.

that procreative decisions can harm or wrong offspring *only* in such cases'.¹⁶ Thus we reach the position that if 'the child has no way to be born or raised free from that harm, a person is not injuring the child by enabling her to be born' *but* this claim 'would not hold if the harmful conditions are such that the very existence of the child is wrong to it'.¹⁷ However, even if the life worth living threshold is accepted and the child's life 'falls below this threshold, then it will again be the case' that the child is not 'any worse off than she could possibly have been' thus necessitating an impersonal account of the harmfulness of bringing them into existence.¹⁸

The terminology *life worth living* can be criticised because 'it excludes the possibility of reasons to carry on living ... which go beyond quality of life' such as when someone 'keeps herself alive for moral reasons, perhaps to benefit third parties' or to see a goal through to the end.¹⁹ While we might agree that there are non-quality related reasons for continuing our existence it is not clear that this would be an applicable consideration in genesis cases as the potential child does not have any goals or third parties it wants to benefit. Does this suggest that the reduction of genesis cases to physiological factors – as suggested by the life worth living threshold – is correct? I think not. For the same reason that non-quality factors must be excluded, physiological considerations must also be excluded for neither can apply to potential persons. The potential person is not an entity that stands in relation to other entities nor to particular conditions – no potential offspring has a reason for benefiting third parties for they do not exist and, equally, they cannot suffer from physiological conditions until they exist. It seems then that this argument regarding non-quality factors in life fails to undermine the life worth living threshold – although it does highlight the reason why both this argument and physiological factors must be excluded from consideration.

However, the non-quality factors are based upon optional goals while physiological conditions are non-optional and are more easily predicted future conditions²⁰. As physiological conditions are scientifically quantifiable the impact of those conditions should also be to a reasonable degree quantifiable. It thus seems that

¹⁶ Ibid 85.

¹⁷ John Robertson, *Children of Choice* (Princeton University Press 1996) 75.

¹⁸ David Boonin, *The Non-Identity Problem and the Ethics of Future People* (Oxford University Press 2014) 64, fn11.

¹⁹ Wilkinson (n1) 70.

²⁰ I am grateful to Prof. Søren Holm for pointing out this distinction to me.

an account based upon physiological conditions can act as a metric for impersonal value after all. But consider the impaired functioning account which seeks to do just this by providing both a minimum positive for the life worth living threshold and a general metric for ranking lives based upon the state of an individual once they exist. Impaired functioning identifies and uses as a benchmark a state of ideal functioning to rank potential (and actual) lives. The problem with impaired functioning accounts is its inherently comparative logic – when we speak of ‘states of impaired functioning ... [as] bad, we are really comparing them unfavourably to states of proper functioning’.²¹ Furthermore impaired functioning presupposes a purpose or particular optimal form of human beings but the variety of human goals should make us cautious in holding that particular functional abilities are necessary for a good life however so characterised.

Any ideal based upon being ‘able-bodied and healthy’ can be criticised on this ground – it requires an objective notion of good functioning which ignores the different lived experiences of different individuals.²² Thus we might exclude physiological conditions, not because their effects are unrealised, but because we cannot judge the impact upon the value of the life lived. It also has about it a sense of human life as purposefully ordered in which humans should be able to perform particular functions or function in particular ways. It is particularly problematic given the normative ideas of ‘healthy’ or ‘normal’ bodies and the status of disabilities, all of which must treat health or normality as fixed and unchanging if they are to be used to judge the conditions that occur in different lives. This means that an impaired functioning account (and judgements concerning physiological conditions) will be contingent on what one believes about the capacities humans should have – itself hardly an objective fact.

Stephen Wilkinson argues that the ‘dividing line between lives containing net *positive* welfare and those containing net *negative* welfare (the ‘zero line’) seems not to be arbitrary’ unlike any super-zero threshold – that is a judgement that a decent, good or acceptable life (however characterised) should be preferred to a *merely* positive life.²³ The zero-threshold standard works as follows:

²¹ Matthew Hanser, ‘The Metaphysics of Harm’ (2008) 77 *Philosophy and Phenomenological Research* 421, 426.

²² Jonathan Glover *Choosing children: Genes, disability and design* (Oxford University Press 2008) 42.

²³ Wilkinson (n1) 76.

[I]f the child could compare the state of affairs in which it exists (one with sub-zero quality of life) with another in which it does not (one with no life, and hence no quality of life) it would (and rationally should), other things being equal, prefer the latter.²⁴

The difficulty with this comparative version is that it requires assigning value to non-existence and different kinds of lives but such comparisons are impossible for ‘there is no way to compare the *amount* of suffering of states of actual people and the state of nonexistence of these people’ as non-existence has no value, no state, no features at all.²⁵ Non-existence is not an alternative condition – it is, by definition, no condition at all. There is the further problem ‘with assuming hypothetical human beings as the subject for whom these ideals are good’ because this misleads us into thinking in person-affecting terms when in fact the decision is not person-affecting and the judgement is impersonal.²⁶ We should, therefore, ‘resist the temptation of assigning a zero-value to non-existence, thus making it quantitatively commensurable’ with the lives of actual people.²⁷ In other words, the existential divide between existing and non-existence cannot be crossed by assigning value to it because non-existence is not a state or condition in which anything exists.

This is shown through the descriptions used – Wilkinson suggests that if the child *could compare* while Buchanan et al say *from the perspective* of someone who will never exist, thus disguising the fact it is not the person whose life it is that is judging the value of their existence, rather it is other agents using a purportedly objective measurement to rank lives that is making the decision. Consequently, the comparative version of a life worth living ‘involves standards requiring comparisons that are metaphysically impossible’.²⁸ Non-existence has no value, negative or positive, and ‘there is no way to compare the *amount* of suffering’ from life with non-existence.²⁹ To suggest that non-existence can be given a status comparable to existence is problematic because the ‘idea of a life having meaning or worth living is

²⁴ Ibid 71.

²⁵ Heyd (n5) 113.

²⁶ Ibid 86.

²⁷ Ibid 113.

²⁸ Ibid 63.

²⁹ Ibid 113.

itself person-affecting ... it can be worth living only *for* its subject' while non-existence cannot have a value for anyone.³⁰

But can existence have positive or negative value? It is clear that we can suffer through existence and it is this that may tempt us to say non-existence has no value (either positive or negative) and to compare it with existence which has positive beneficial value.³¹ In other words, it is the suffering inherent in existence that may tempt us to say that non-existence as the absence of suffering has comparative value. In contrast, David Benatar holds the view that existence has negative value because of all the harm and suffering that life brings with it.³² Thus existence is itself harmful rather than beneficial and, if correct, bringing someone (anyone) into existence is harmful. Benatar's view is firmly based on the effects of life on those who exist – '[n]one of this [suffering] befalls the non-existent. Only existers suffer harm'.³³ However, this conclusion rests upon the claim that 'the absence of pain is good, even if that good is not enjoyed by anyone' while 'the absence of pleasure is not bad unless there is somebody for whom this absence is a deprivation'.³⁴ Benatar seems to rely on an impersonal account of pain (for none need benefit from the absence of pain) against a person-affecting notion of pleasure (pleasure must be felt by someone) which reflects 'the commonly held view that ... there is a duty to avoid bringing suffering people into existence'.³⁵ (It is not clear what he means by suffering people – on the basis of the asymmetrical value of pain and pleasure it would seem that all people are suffering people and so reproduction should not occur at all.)

Benatar argues that this reflects the 'commonly held view' that we have 'no ... positive duties to bring about happiness' and that 'sometimes we do avoid bringing a child into existence because of the potential child's interests'.³⁶ This, however, is to assume what is under discussion – that is Benatar seems to assume that bringing someone into existence is harmful and that the asymmetry holds. However, it is unclear why a person-affecting view applies to pleasure but an impersonal view applies to pain when they are both subjective states of agents (and other living things).

³⁰ Ibid 110.

³¹ David Benatar, 'Why it is Better Never to Come into Existence' (1997) 34 *American Philosophical Quarterly* 345.

³² Ibid 345.

³³ Ibid 345.

³⁴ Ibid 346.

³⁵ Ibid 346.

³⁶ Ibid 346.

For the moment let us take Benatar's view as given – what are the implications of his view that we should prevent suffering but are not obliged to produce happiness? It suggests that we should not bring any children into existence at all because that will create suffering which we have a duty to prevent. Therefore we can be morally reprimanded for all the suffering we create but we will not be reprimanded for the all the happiness we fail to produce. If this view were transcribed into legal regulation it would lead to the prohibition on accessing ART entirely – whatever kind of child prospective parents would have. Of course we might not subordinate the preferences of prospective parents to this particular moral claim but if we accepted that Benatar is correct (which we can justifiably be sceptical of given the asymmetrical basis of his claim) it would seem that all those we create would have a life of negative value and so every child produced would be harmed by their parents.

While Benatar's claim is suspect, it does demonstrate the need for an impersonal basis to judge the value of a potential life – this may be why Benatar retains the asymmetry. Only by using objective mind-independent impersonal harm we can say that one life is more worthwhile than another but this permits the insertion of any particular value that we subjectively hold and argue that this should form the basis of ranking lives.

The Fetishism of Impersonalism

The difficulty in establishing a criterion to measure the worth of life seems to be due to impersonalism being 'an abstract concept that is difficult to pin down, analyse and thus criticize' but this also gives it the flexibility to reinforce different intuitions of harm.³⁷ We can see this from the way in which arguments regarding better children or more worthwhile lives are framed – an appeal can be made to whatever someone thinks is good, valuable or important in life. Consider Bonnie Steinbock's notion of a decent life. The decent minimum standard (DMS) holds that for 'life to be a positive benefit, certain minimal conditions must be satisfied', which are 'reached only if life holds a reasonable promise of containing the things that make human lives good'.³⁸ This threshold includes such things as 'an ability to experience pleasure, to learn, to have relationships with others' and it imposes an obligation upon prospective parents

³⁷ Bennett (n9) 266.

³⁸ Bonnie Steinbock, 'Wrongful Life and Procreative Decisions' in Melinda Roberts & David Wasserman (eds), *Harming Future Persons: Ethics, Genetics and the Nonidentity Problem* (Springer 2009) 163.

to have ‘the ability to be a good enough parent[s]’.³⁹ All of these requirements are highly subjective – not everyone seeks relationships with others, not everyone holds learning as key part of their life and what counts as being good enough parents will be so subjective as to make the phrase almost meaningless – but they are portrayed as the objective criteria of lives and parenthood.

Thus any normative claim can be given an impersonal form and used to rank lives – for example, it seems fairly common for people to hold non-disability as a *good* state of existence and so impersonal claims framed on this ground will intuitively appeal to a many. Thus it seems that ‘the plausibility of the this [sic] notion of impersonal harm stems solely from its ability to explain the intuitive response’ we have to certain reproductive choices.⁴⁰ Just like the concept of the *good life* any belief can be used as the basis of a purportedly objective value of life – but this simply raises the problems of cognitivist morality shown by moral error theory even if we get past the subjectivity of the claims involved. The flexibility of impersonalism in appealing to existing intuitions explains the tenacious use of impersonalism to justify particular normative positions. But this is the fundamental problem with an impersonal approach – it ignores the subjectivity of those whose lives it purports to rank allowing the imposition of whatever a particular exponent considers as the most important features of life.

A striking example of impersonalism being used to reinforce existing normative beliefs is provided by Buchanan et al’s rejection of a person-affecting account because ‘the principles of traditional ethical theories apparently fail to support [the] ... judgement’ that the young girl is wrong to go ahead with the pregnancy.⁴¹ They propose as an alternative, the impersonal substitution principle, to cover genesis cases. The substitution principle is as follows:

Individuals are morally required not to let any child ... for whose welfare they are responsible experience serious suffering or limited opportunity⁴² or serious loss of happiness or good, if they can act so that, without

³⁹ Ibid 163: We will not take up the role of parental obligations upon prospective parents here.

⁴⁰ Bennett (n9) 268.

⁴¹ Buchanan et al (n15) 247.

⁴² This term is objectionable on the basis everyone will experience a great deal of limited opportunity in some avenues of life while experiencing opportunities in other areas. Presumably the term here is meant to convey the idea of barely any or no opportunities in life but this would have to cover cases such as children born into poverty and perhaps would also require a higher than worthwhile threshold.

affecting the number of persons who will exist ... no child ... will experience serious suffering.⁴³

Consequently, while ‘suffering and limited opportunity must, of course, be experienced by some person’ – and so in this sense ‘remains person-affecting’ – this principle ‘does not require that the same individuals ... experience suffering’: ‘it is a same number, not a same person, principle’.⁴⁴

Use of this principle allows us to maintain that while the ‘wrongness [in genesis cases] cannot be explained in terms of harming or wronging or making an actual future person worse off’, ‘[n]evertheless, we can explain the wrongness by appealing to an impersonal substitution principle’.⁴⁵ Applying the substitution principle to the young girl would require that she waits and has the second possible child whose life will contain less suffering than the child she would have had at the age of fourteen (assuming that more suffering in present in the first child’s life). But the adoption of the substitution principle is based upon the refusal to accept the conclusion that no harm is caused to offspring. There are two problems with the move to a substitution principle: first, it fails to be independent of the desired solution and, second, the implications of an impersonal approach cannot be restricted to potential agents but must also include actual agents because impersonalism is a fundamental axiological position.

Regarding independence, David Boonin argues that ‘the reason provided for rejecting a given premise of the non-identity argument must be independent of the fact that rejecting the premise would enable us to avoid the Implausible Conclusion’ (Boonin’s terms for the person-affecting non-identity conclusion).⁴⁶ In the case of the substitution principle, which rejects the premise that if ‘an act does not harm anyone, then the act does not wrong anyone’, the only justification given by the above claim seems to be that it allows us to reject the non-identity conclusion.⁴⁷ The only reason to accept the substitution principle seems to be because it allows us to ‘explain the wrongness [of genesis cases] by appealing to an impersonal substitution principle’.⁴⁸

⁴³ Buchanan et al (n15) 249.

⁴⁴ Ibid 249.

⁴⁵ Ibid 91.

⁴⁶ Boonin (n18) 20.

⁴⁷ Ibid 5.

⁴⁸ Bonnie Steinbock, *Life Before Birth* (2nd edn, Oxford University Press 2011) 91.

Further complicating matters, Buchanan et al state that they do not reject person-affecting principles but hold that ‘an adequate moral theory should include as well non-person-affecting principles’⁴⁹. The reason given for this, as far as can be inferred from their book, is that we should ‘abandon the specific feature of typical moral principles about obligations to prevent or not cause harm which generates difficulty ... in genesis cases’ – namely the ‘person-affecting property of principles beneficence and nonmaleficence’.⁵⁰ But they provide no explanation of how both person-affecting and impersonal harm can operate together – even if we rejected the claim that ‘objectivity in Ethics must be all-or-nothing’.⁵¹ It is difficult to see how this would work:

[We] cannot eat the cake of attributing utility only to those who can be said to better or worsen *their* lives and have the cake of a global preference for a world with more happiness for certain people over a world with less happiness for different people.⁵²

We cannot have both systems because there is no way to reconcile them in cases of conflict. One refers to properties independent of agents while the other relies on agents for determining value but they cannot operate in unison.

If these are the only reasons given for adopting an impersonal principle of harm, we can (indeed must) reject this move for failing the independence test. The only reason given for proposing the impersonal solution is to avoid the conclusion that we cannot harm potential persons. Buchanan et al’s only reply is that it does cause harm on an impersonal conception of harm which is no answer at all. The important question is whether we should adopt a person-affecting notion of harm or an impersonal one but this must be determined independent of the conclusion that we cannot harm those who we create.

Derek Parfit suggests that we adopt impersonalism on the basis ‘that one person can be worse off than another, in morally significant ways, and by more or less’.⁵³ Moreover, Parfit accepts that if we want to claim that children can be harmed

⁴⁹ Buchanan et al (n15) 250.

⁵⁰ Ibid 248.

⁵¹ Derek Parfit, *Reasons and Persons* (Oxford University Press 1984) 452.

⁵² Heyd (n5) 89.

⁵³ Parfit 357.

by the action that creates them impersonalism ‘is what we *must* claim’ because as we have seen person-affecting harm cannot function in genesis cases.⁵⁴ Parfit argues for impersonalism as an axiological position thus making the impersonal conclusion independent of the desired answer. However, while he supports impersonalism he is discomforted by the Repugnant Conclusion which concludes that a sufficiently large population of people with very low amounts of impersonal value (e.g. well-being, health or whatever) produces a greater total value (whether based on average or total amounts) than a smaller population with individually higher amounts of value.⁵⁵ Impersonalism is ‘a principle about the quality and quantity of the lives that are lived, but is not about what is good or bad for those people whom our acts affect’.⁵⁶ It therefore explicitly ‘ignores the boundaries between lives, the separateness of persons’ and thus colonises agents by homogenising individual agents into interchangeable moral beings and producers of value.⁵⁷ Because of the Repugnant Conclusion Parfit rejects his model of impersonal value, arguing that we need to ‘find a better version: Theory X’ which would take the form of a system of impersonalism that *does not* lead to the Repugnant Conclusion.⁵⁸

But the Repugnant Conclusion is not a reason for rejecting impersonalism – it is the inevitable outcome of treating individuals as producers of (whatever) metric of impersonal value that is used. It is an inevitable outcome because any account that treats individuals as producers of a universal utility uses a calculus that leads to a sufficiently large population of individually impoverished agents to produce more value than a smaller population of individually higher valued agents. Parfit rejects the Repugnant Conclusion not because it is inconsistent with his theory of impersonalism but because it ‘is *intrinsically* repugnant’.⁵⁹ On an impersonal account the outcome is the morally better outcome and Parfit rejection is seemingly based on intuitionism – the Repugnant Conclusion strikes us as abhorrent and so it must be wrong. If one truly subscribed to the impersonal perspective the conclusion should be that our reaction is wrong while the so-called repugnant outcome is morally preferable. The Repugnant Conclusion is the logical outcome of impersonalism and Parfit’s Theory X little more

⁵⁴ Ibid 360. My italics.

⁵⁵ Ibid 381-90.

⁵⁶ Ibid 446.

⁵⁷ Ibid 446.

⁵⁸ Ibid 390.

⁵⁹ Ibid 390. Original italics.

than an attempt to avoid that repugnant outcome – an outcome he seems to reject for intuitionist reasons alone.

Person-affecting and Non-person-affecting Harm

We can see that person-affecting and impersonal normative claims are incompatible. While person-affecting harm can be empirically verified by our own experiences (given a clear definition of person-affecting harm) impersonalism cannot requiring cognitivist moral claims in order to function. Moral error theory gives us reason to be highly suspicious of impersonal claims to harm because they require belief in irreducibly normative principles which any agent can reject. On the basis of person-affecting harm the welfare of future people cannot justify the constraints imposed on prospective parents accessing ART. On a person-affecting account genesis choices are ‘outside the realms of morality: it is a morally neutral choice, a preference’ held by existing agents about the reproductive choices of other agents.⁶⁰ (Preferences about the preferences of other agents cannot be enforced because this will result in instant conflict as each agent tries to maintain the integrity of their preferences.) The system of constraint imposed by the regulation of reproduction can be justified when harm occurs to others (which does not include harming preferences about the kinds of children other people should have, rather it is about protecting agents themselves from interference).

Fundamentally, the conflict between person-affecting harm and impersonal harm is an axiological conflict between ‘certain absolute characteristics of the world as against certain person-relative states of the world’.⁶¹ The question is ‘whether value (in general, but also moral value in particular) is an “impersonal” attribute of the universe’.⁶² However, person-affecting harm is a superior model of reality because it occurs whether or not someone has a belief or acceptance in such harm – if someone stabs me that is a harm whether I believe in impersonal harm or not. Thus person-affecting harm cannot be excluded because our actions do affect others and will provoke reactions – this after all is how interaction works.

The alternative is to embrace impersonalism but that requires metaphysically dubious cognitivist moral claims which can be rejected (as argued in chapter one)

⁶⁰ Bennett (n9) 269.

⁶¹ Heyd (n5) 80.

⁶² Ibid 81.

simply by assertion. Additionally, we have reason to doubt that impersonal goods and harms are plausible because the good which forms the basis of impersonalism (whatever it is) is ‘by necessity *of actual*’ agents thus presupposing the existence of agents and – so far at least – takes the form of a personal preference.⁶³ Take for instance, the claim that a life is not worth living from the point of view of the potential person – this future person is clearly not making a judgement about their life, we are imagining how they would behave using our judgement of their *interests* or using ideals of functioning, good or decent lives. But ‘interests and even ideals are by necessity *of actual human beings*’.⁶⁴ Aside from the conceptual difficulties impersonalism is problematic because it ‘sacrifices the utility of individuals to the promotion of the impersonal value of the overall good’ thus failing to provide a reason *for agents* to bring about those outcomes.⁶⁵ Indeed this may be the root of Parfit’s problem with the Repugnant Conclusion because it allows abysmally miserable lives to be produced on an enormous scale but if individuals are important in their own right this is a problem.

However, because impersonalism entails a metaphysical commitment to cognitivist moral claims it is not truth-apt – consequently, if someone believes or accepts impersonalism this is not, ultimately, a perspective that can be invalidated (although I believe I have set out a number of reasons to reject impersonalism). If it comes down to belief and assertion then impersonalism cannot be wholly discounted. Nevertheless impersonalism cannot justify population wide constraint for that very reason – any agent can reject it in favour of a person-affecting account of harm. Thus while regulation can be based upon person-affecting harm (for this is empirically true of human existence) we cannot ban those who adopt an impersonal view from acting according to their own beliefs. Fortunately, in the regulation of access to ART, person-affecting based regulation (along the lines set out here) permits prospective parents to access ART whether they accept impersonalism or not. In other words, parents can have the children they wish – whatever kind of child that might be – without prejudice. Such a position is not possible if regulation is based upon impersonal harm or based upon the mistaken application of person-affecting harm to genesis cases.

⁶³ Ibid 63.

⁶⁴ Heyd (n5) 63.

⁶⁵ Ibid 57.

We therefore have a number of reasons for rejecting impersonal harm. First, it relies on cognitivist moral claims concerning the metric by which lives are measured and ranked which means that such objective claims are not truth-apt. This metric must be irreducibly normative if it is to apply to potential persons for it cannot be based upon the effect on the not yet existent individual – rather increasing the amount of, say, health must be in and of itself something we should do. Consequently, impersonalism is subject to the critique of moral error theory and can be rejected on the basis that impersonalism can only be asserted. There is a further problem with the use objective criteria in impersonalism for it allows different authors to use their own subjective preferences and coat them in a veneer of objectivity – as we saw with Savulescu’s work, Buchanan et al’s substitution principle, the notion of impaired functioning and the decent minimum standard.

Even if a conclusion objective metric were discovered it would still rely on homogenising all agents into a numerical mass ignoring the importance of individuals in determining their own lives. Thus a couple who wanted a disabled child would not have any reason to comply with an impersonal metric that states they are producing a less worthwhile child. Finally, while impersonalism can only be asserted – so that no one can be required abide by its strictures – this does not apply to actions affecting extant agents. Those who subscribe to an impersonalist account should not, however, be prevented from acting according to that system because it concerns their beliefs about what their lives should be like. But for legal regulation that applies to the entire population of agents impersonalism cannot justify constraint. We must therefore base legislation on person-affecting harm and, consequently, genesis cases are excluded from judgements of harm or benefice because our action do not affect a subject but create one.

Chapter 5

*Applying the Actual/Potential Person Distinction to Reproductive Tort*⁶⁶

(This chapter has been published as Samuel Walker ‘Applying the actual/potential person distinction to reproductive torts’ (2014) *Medical Law International* as it appears here, except for the formatting which has been adjusted to match the rest of this thesis)

Abstract

As technology has advanced the level of control that can be exercised over the reproductive process has increased. These advances have resulted in a number of claims in tort law relating to pregnancy and birth. The three reproductive torts considered here are ‘wrongful conception’, ‘wrongful birth’ and ‘wrongful life’. This paper will consider the theoretical underpinnings upon which these torts rest, and will suggest that the potential/actual person distinction is crucial to these reproductive torts because potential persons should not be able to make claims in tort based on alternative conditions that could never have been. This is because actions (or omissions) prior to birth determine the pre-conditions for existence. Thus, only actual persons (that is those who exist at the time of the action or omission) should be able to bring claims in tort. The analysis will conclude by arguing that no child should be permitted to bring a claim under any form of reproductive tort.

Introduction

The main purpose of this work is to consider the implications of acknowledging the potential/actual person distinction in relation to legal claims relating to reproductive torts. This entails consideration of what constitutes a ‘person’ and what impact decisions in this area will have upon the existence and identity of potential persons. The discussion in this paper will be anchored around a discussion of the reproductive torts of wrongful birth/conception and wrongful life claims. For the sake of simplicity,

⁶⁶ The term Reproductive torts originates with Nicolette Priaux’s work and encompasses all three terms; wrongful life, wrongful birth and wrongful conception. While these are distinct terms, all fall within negligence for the purposes of this analysis ‘reproductive torts’ is a useful term to identify these particular claims: Nicolette Priaux, ‘Health, Disability & Parental Interests: Adopting a Contextual Approach in the Reproductive Torts’ (2005) 12 *European Journal of Health Law* 213.

wrongful birth and wrongful conception will be discussed together because they both relate to parental claims rather than children claiming. These torts have been selected because the very basis of these claims relates to the existence and identity of a person, *who would not have existed but for the action giving rise to a tort claim*. Collectively these torts deal with actions that cause an individual to exist. Furthermore the reproductive torts of wrongful conception, birth and life are pertinent to discussing liability for reproductive actions because they intersect with claims between parents, offspring and third-party medical providers. Reproductive torts thus provide a suitable context within which to discuss the potential/actual distinction, psychological personhood and the implications for legal reform.

The format of this article will be as follows. Firstly it will provide an outline of why the potential/actual person distinction is important and what that implies for how we think of decisions that create new individuals. This will be followed by a section assessing the construct of psychological personhood, which is important because it highlights that changes occurring *before* a potential person becomes an actual person creates new individuals rather than changing the conditions for a pre-existing one. This argument is based upon the fact that different psychological characteristics create different people and that this is affected by both mental and physical changes to an organism. Thus, prior to actual existence no person exists and changes to the conditions in which offspring will be produced are creating new persons. Together these two related concepts of potential/actual persons and psychological personhood will serve as a platform for reassessing reproductive torts and in suggesting the direction reform of the law should take. Essentially the claims will be as follows; potential persons have no claims on the actions of actual persons and only the interests of actual persons can sustain claims in tort.

As we consider the reproductive torts of wrongful birth/conception and wrongful life we will see how the courts have encountered the problems which the theoretical section on potential/actual persons and psychological personhood expose. In further support of the integral nature of the problems with reproductive torts, other non-UK jurisdictions will be brought into the discussion. For the sake of clarity, the differences between the reproductive torts under consideration will be explored in the brief description of each 'tort' that follows.

‘Wrongful conception’ - Wrongful conception cases concern the birth of a healthy child where the very conception of a child was something that the parents had actively sought to prevent (for example by undergoing a vasectomy which was carried out incorrectly, so that the parents now have a child which they never intended to have.⁶⁷ ‘Wrongful birth’ cases occur when a child has been conceived and the parents are given incorrect advice or information regarding the condition of the foetus, where such advice if given non-negligently might have led to a termination.⁶⁸ These cases also include situations where some action by the medical professional causes a disability to occur or damages the foetus.⁶⁹ Normally these reproductive torts are relatively straight forward claims by existing persons (the parents), that they have been harmed because the medical procedure was carried out incorrectly and resulted in the birth of a child, with or without a disability.

However in England and Wales⁷⁰ there is an aberrant situation under s.1 of the Congenital Disabilities (Civil Liability) Act 1976 (‘CDCLA’) which allows a child to claim under the tort of wrongful birth when the actions of a medical provider lead to the child having disabilities that it otherwise would not have. It is important to note that this does not allow claims by healthy children; it only allows claims under wrongful birth in cases of disability caused by medical providers. In this respect wrongful birth claims under the CDCLA are somewhat similar to the wrongful life cases.

‘Wrongful life’ - The tort of wrongful life is a claim by a child that they should not have been caused to exist at all. In essence offspring claim that life is so intolerable and full of inescapable suffering that *they are better off not existing* rather than being born in such a state. The courts are thus asked to conclude that being born was harmful to the child and that non-existence was a better condition, of which they have been deprived. The term wrongful life is not without controversy as Kirby J in the Australian case of *Harriton v Stephens* demonstrates. Kirby J stated that in his view ‘its use, even as a shorthand phrase, should be avoided’.⁷¹ Kirby J takes this view for a number of reasons, but the most important of them are as follows; the

⁶⁷ *MacFarlane v Tayside Healthboard* [2000] 2 AC 59.

⁶⁸ *Farraj v King’s Healthcare NHS Trust* [2010] 1 W.L.R. 2139.

⁶⁹ *Whitehouse v Jordan* [1981] W.L.R. 246.

⁷⁰ This has been corrected from the published version which uses ‘UK’.

⁷¹ *Harriton v Stephens* [2006] Unreported Cases High Court of Australia [8].

phrase wrongful life ‘denigrate[s] the value of human existence’ and that by ‘lumping all such cases under the one description there is a danger that important factual distinctions will be overlooked or obscured’.⁷² Kirby J prefers the term ‘wrongful suffering’ as it encapsulates, what Kirby J sees as the basis of the claim, that the ‘negligence ... has directly resulted in the present suffering’.⁷³ In line with judicial terminology the phrase wrongful life will be retained in this article, although as we will see, in addition to the just stated objection, this term has different interpretations in different jurisdictions.

However, this raises the philosophical problem of comparing existence with non-existence which is conceptually impossible because non-existence is the absence of everything. There is thus *no-one* who is better off than the child that does exist. This is also a legal problem for the courts as they cannot (and are not prepared to) calculate damages on this basis and are unwilling to countenance the idea that non-existence could be preferable to life⁷⁴. Currently, such claims are not permitted under English and Welsh law⁷⁵ and, as we shall see, other jurisdictions have struggled with the wrongful life tort. In the New Jersey case of *Gleitman v Cosgrove* [1967]⁷⁶ the court found that a wrongful life claim could not be actionable, while in the state of California the case of *Curlender v Biosciences Laboratories* [1980]⁷⁷ held that a wrongful life claim was sustainable. It is thus by no means a settled question as to whether wrongful life claims can be judicially recognised.

Moreover, wrongful birth and wrongful life actions can be distinguished as claims brought by parents and claims brought by the offspring themselves respectively, however this division on the basis of the identity of the claimant does not always apply, as we have seen from the CDCLA.⁷⁸ Throughout this article, these reproductive torts are separated by type (that is wrongful birth/conception and wrongful life) rather than on the basis of who brings the claim. It should, however, be borne in mind that who can claim, normally, follows the pattern of ‘wrongful conception/birth’ claimants being the parents, while ‘wrongful life’ claimants are the

⁷² Ibid [11].

⁷³ Ibid [10].

⁷⁴ *McKay v Essex Area Health Authority* [1982] Q.B. 1166.

⁷⁵ Ibid. (This has been corrected from the published version which uses ‘UK law’).

⁷⁶ *Gleitman v Cosgrove* [1967] 49 N.J. 22.

⁷⁷ *Curlender v Biosciences Laboratories* [1980] 165 Cal. Rptr. 477.

⁷⁸ *Harriton* (n6) [11].

children (or someone acting on behalf of the child); but this is not always the case and sometimes the claimants will be different as under the CDCLA.

Reproductive torts relate to harms and damages that the law recognises to offspring where the action has occurred before their birth, and covers both parents and the child claiming against a medical provider. The law can therefore recognise that harms can occur before a legal person comes into existence and that once the child exists their claim crystallises allowing a claim in tort. However the conclusion of the paper will be that children can never claim for the condition they are born in, regardless of what state that is or how it was caused. This would require the removal of section one of the Congenital Disabilities Act which allows children to claim under wrongful birth cases, and retain the non-recognition of wrongful life torts. This conclusion will be based upon the fact that decisions and conditions which constitute the new person cannot be grounds for claims that they either should have existed in a different condition or should not exist at all, because in both cases they (the person claiming) would not exist and thus they would not be affected. This line of reasoning is ‘person-affecting’ in its approach, which dovetails with tort law as tort requires that *a person* is in a worse condition than they otherwise would be; that is a person is affected by the wrongful conduct.⁷⁹

This conclusion would apply whether the conduct comes from medical providers, prospective parents or genetic engineers (whether through genetic selection or modification) thus only allowing for the interests of actual persons to determine the decisions made. Therefore the legal claims of reproductive torts will not be able to be based upon harm to the not-yet-existing offspring but would have to be based upon the effects of their actions on actual persons.

The Potential/Actual Distinction⁸⁰ and Psychological Personhood

The importance of the potential/actual person distinction cannot be overstated because it speaks to the difference between those who can suffer harm or gain some benefit and those whose very existence is constituted by the decisions being undertaken. It is important because the basic aim of tort law, including reproductive torts, is to put someone in the situation they would have been in if the tortious action had not

⁷⁹ Catherine Elliott and Francis Quinn *Tort Law* (Longman, 2011) 375.

⁸⁰ The potential/actual distinction is a concept that arises across the literature. See, for example, David Heyd, *Genethics* (University of California Press 1992) and Alberto Giubilini and Francesca Minerva ‘After-birth abortion: Why the baby should live?’ (2012) 39 *Journal of Medical Ethics* 261.

occurred⁸¹. The problem with reproductive torts is that the situation before the tort occurred is non-existence, and thus a comparison between existence and non-existence is required to calculate damages. As Ackner LJ states ‘how can the court begin to evaluate non-existence ... No comparison is possible and therefore no damage can be established which a court could recognise’ (*MacKay v Essex Area Healthboard* [1982] at p.1189). Thus we do not have to consider whether we take a person-affecting or impersonal approach to harm, because tort presupposes that persons *are* affected. We therefore must be able to identify who the person affected is and we must know how they were affected and what condition they would have been in prior to the action.

We can thus see why the potential/actual person distinction is important. Decisions which create persons cannot be subject to considerations of harm and benefit in relation to the person yet to exist, because the very conditions for affecting persons require that they actually exist and potential persons do not fulfil this requirement. Moreover if, as in the case of reproductive torts, the action which is complained of did not occur then the person would never have existed and thus there would be no ‘person’ in a better or worse condition. As David Heyd puts it, the ‘confusion created by wrongful life cases arises precisely from the fact that the wrongful act is the direct cause of the plaintiff’s existence’.⁸² Although Heyd is speaking only in relation to the tort of wrongful life, this paper will argue that the person-affecting requirement of tort is never satisfied in reproductive cases when the child brings the claim. This is because potential persons do not exist, and as such changes to the embryo’s genetic make up or physical body do not change *anyone*; rather they alter the genesis conditions which create a person. For example, changing whether an embryo has a disability will not change the (potential) person who will exist because no one exists, it will only change the conditions of the creation process.

Additionally the existential status of potential persons means that they also fail to fulfil the need for a single relationship. As Ernest Weinrib states tort law, as a component of corrective justice, ‘treats the doer and the sufferer of harm as the active and passive participants in a single relationship’.⁸³ Thus the legal relationship required

⁸¹ Elliott and Quinn 375.

⁸² Heyd (n15) 29.

⁸³ Ernest Weinrib ‘The Jurisprudence of Legal Formalism’ (1993) 16 *Harvard Journal of Law and Public Policy* 583, 588.

in tort claims is absent because of the necessary requirement of existence for there to be a doer and sufferer, a harm and a harmed.

What are Psychological Persons?

This paper adopts a psychological personhood approach to the nature of persons and personal identity which will now be outlined. The concept of psychological personhood is a claim that can be summarised as; persons exist only as an aggregate or a construct of the mental capacities of organisms, they do not exist outside of this realm. Thus a person is not a person because they are biologically human, or because they possess a name, but because they have certain psychological features of memory, self-awareness, a sense of self over time, desires and interests between the past, the ongoing now and future instants.⁸⁴ These features then give rise to psychological *connectedness* (connections between one instant and the next) and together form psychological *continuity* between past, present and future.⁸⁵ Lynne Baker identifies the crucial attribute of a person as the ‘First-Person Perspective’ (FPP) in which a being can think of itself as the source of thoughts and desires.⁸⁶ This is a good means of identifying a person and provides a continuous sense of identity within the entity that is a person. The content of personal identity, that is, of a person being a particular someone and having specific goals and desires is provided by the other psychological features of experience and memory. Thus the FPP provides a sense of unity over life and experience, but FPP relies on connections and continuity between instants and experiences to provide personal identity. Personal identity is therefore the narrative of a specific person’s existence, while being a member of the group of beings called ‘person’ is dependent on the capacity for a First-Person Perspective. If these connections are absent, then psychological personhood would suggest that someone is not the same person over time even if they remain the same physical organism. This allows for a self-referencing narrative which builds a biographical life, gives meaning and purpose of a person’s life and makes a life plan constructed through projects⁸⁷ possible; this self-recursive construction is what distinguishes a person from a non-person.

⁸⁴ John Locke, *An essay on human understanding* (first published 1689, Everyman 1996); Derek Parfit, *Reasons and Persons* (Clarendon Press 1984) 204-9.

⁸⁵ Parfit 204-9.

⁸⁶ Lynne Baker, *Persons and Bodies* (Cambridge University Press 2000) 64.

⁸⁷ Loren Lomasky, *Persons, Rights and the Moral Community* (Oxford University Press 1990) 32.

This starkly contrasts with how we normally think of ourselves. As a consequence, the importance of psychological attributes is greater than under alternative theories of persons (such as the view of persons being an indestructible soul or being a human organism). Furthermore, it means that persons are fundamentally the font of value and for things to be ‘good’ or ‘bad,’ a person must *actually* exist. An actual person who exists can be affected by things that happen around and to them, and thus can make a value judgement regarding those actions. Potential persons cannot be affected in a similar manner because the necessary condition for them to evaluate the effect of things is that they exist, but it is their very existence which is being decided.

Thus Heyd can state that ‘they [potential persons] do not meet certain preconditions of existence and identity’, and consequently as we modify and change an embryo (that is, as we change the conditions for the potential person’s existence) we will be creating a new person, not altering the circumstances for the same person.⁸⁸ If this were not the case, for example, if person were ‘souls’, then reproductive torts would be sustainable, as the existence of the person would be prior to the existence of their body. However, in this case, every negative attribute would be included in a tort claim because these would all be *affecting someone* and would be making things worse for them. In the case of psychological persons, the person develops after their body, thus actions which occur prior to their existence are excluded from *affecting* them because these actions constitute the person who will exist.

This theoretical platform means that actions giving rise to reproductive torts will lead to a new person coming into existence regardless of what these actions are. Thus medical providers causing a disability to occur in an embryo will simply be changing the constituting conditions of a new potential person. Similarly wrongful conception results in a new potential person who would not otherwise exist. Equally, failing to diagnose a disability that would lead to the parents terminating the pregnancy would also be creating a new person. In both scenarios, no-one is being made better or worse off, no-one is being replaced with a different person; rather each scenario creates a new person, generating a new personal identity from the specific constituting conditions. As wrongful life claims would have seen an actual person not

⁸⁸ Heyd (n15) 36.

come to exist, potential persons have no claim upon actual persons and wrongful life claims must be rejected. But we will now consider opposing views to this theory.

Potential Persons and the ‘Right’ to be Born (Healthy)

Having set out the argument for reforming reproductive torts by eliminating claims made by the child themselves, it is only fair that some counter claims arguing that potential persons have interests are considered. The obstacle that must be overcome in order for potential persons to have any interests or rights over actual persons is the potential/actual distinction, because it is this divide that precludes claims based upon actions that affect the constituting conditions of a person’s existence.

So how then can interests and rights translate across the potential/actual divide? Loren Lomasky suggests that harms that affect a future person, but which are caused before the person exists can give rise to a retroactive complaint. This is because actual persons can recognise ‘goods-for-the-child [or potential person] that others have reason to acknowledge and respect’, and recognise that the future person is ‘identifiable as a distinct individual upon whom one can act for better or ill’.⁸⁹ Walter Glannon suggests that we can ‘prevent actual (future) people from experiencing pain and suffering and thereby avoid defeating their interests in having healthy lives’.⁹⁰ Moreover, Jonathan Glover states that ‘the parental desire for a healthy child fits with the interests of the child born as a result of the choice’.⁹¹ These statements suggest that we can both identify the person who will exist and extrapolate from that things which are/will be good for them.

Lomasky claims that recognising ‘good-for-someone’ and being able to identify them provides a way to identify harms (over which we have control) for future persons thus we should act to prevent these harms occurring. In other words, because offspring will become persons *who are now identifiable* and we can predict the impact of our actions on this future person we are subject to their interests. Similarly Glover and Glannon claim that a potential person has an interest in being born in a healthy condition. It thus seems that, to Glover at least, once a conception occurs we can identify some interests of the future person. Glover names, for

⁸⁹ Lomasky (n19) 161.

⁹⁰ Walter Glannon, ‘Genes, Embryos and Future People’ (2002) 12 *Bioethics* 187, 193.

⁹¹ Jonathan Glover, *Choosing children: Genes, disability and design* (Oxford University Press 2008) 42.

example, an interest in being ‘able-bodied and healthy’.⁹² Therefore ‘the interests of the child should set limits to what potential parents do’.⁹³

Glover’s approach seems to satisfy Lomasky’s identifiable individual requirement, it therefore seems that a potential person can in some sense be an identifiable subject. This is, however, an error because potential persons are not subjects; that is, they (a potential person) cannot be affected by actions which will constitute the person’s conditions of existence. When we consider how these changes would be brought about we can see the claims above do not work. If we cannot change the ‘health’ of an embryo, then the potential person can only exist in an ‘unhealthy’ condition. If we can change an embryo to be healthy then it would change the constituting conditions of the person being created and thus would not be affecting them because we are creating a new person, making the ‘identifiable’ individual no longer identifiable. Thus ‘we can affirm that people cannot logically have a right to any genetic endowment, if that constitutes their identity’.⁹⁴ Health, amongst many other (and possibly all) attributes, changes the conditions constituting the potential person, such that each change is a fresh act of creation.

It could be claimed that embryos do have interests which only crystallise once they become a person. Yet this makes no logical sense for the same reason that both Glover and Glannon’s ‘interest’ and Lomasky’s ‘good-for’ approaches cannot work; because each approach requires a subject who benefits from the good or interest. However, the persons who supposedly possess these interests or for whom the goods are good-for are constituted by such decisions; that is, the conditions for being benefited or harmed require a person, a subject to exist, in order to determine value to that person. Just as there is no basis for saying that being brought into existence is a benefit, because no subject exists who is waiting to be alive, it cannot be claimed that a potential person is benefited by being born in a particular condition. And ‘for the same reasons, we cannot hold the child is ... an object of maleficence’; in other words, a potential person who becomes an actual person cannot be benefited or harmed by decisions which create and cause them to exist.⁹⁵ There are no exceptions to the statement that potential persons cannot be harmed by being caused to exist;

⁹² Ibid 42.

⁹³ Ibid 43.

⁹⁴ Heyd (n15) 172.

⁹⁵ Ibid 109.

only actual persons have interests which can be harmed and thus any regulation has to be based, and can only be based, upon the interests of extant persons.

The Importance of Reproductive Torts

We can now turn to the changes and impact that reproductive torts would have to undergo in order to conform to the idea that potential persons have no claim over actual persons. Allowing parents (actually existing persons) to claim creates no difficulty at all because they have interests that can be affected by the actions which lead to the existence of offspring. On the other hand, allowing children who result from actions which create them to claim against someone for acts or omissions causing their existence causes a great deal of difficulty. Reproductive torts would need to be reformed because it is logically impossible for a child to be harmed or benefited by actions which create them. Furthermore as we have seen the necessary relationship between the person who causes the harm and the person who suffers the harm is absent, because both parties must be in a ‘single [correlative] relationship’,⁹⁶ It is thus unjust for medical providers to be held as having harmed someone when this cannot be the case. Moreover, if it is the action which created the individual that is harmful then this applies regardless of whose action it is; thus if parents decide not to abort or prevent conception (where they have had accurate medical advice) then they are as much to blame as medical providers and should be liable under reproductive torts. As we will see when considering the United States case of *Curlender v Bio-Science Laboratories*⁹⁷, the logical permissibility of a child suing their parents has been raised in court before. As Rosamund Scott has highlighted however, allowing offspring to claim against their parents would be problematic, not only because of the intra-family conflict that might arise, but because it will also ‘come up against a woman’s very personal moral interests in self-determination’.⁹⁸ This of course is not a bar to the possibility of allowing offspring to claim against their parents, but explains why we should avoid it.

Even with these concerns put to one side, reproductive torts still face great difficulty in terms of agreeing a basis for calculating damages. As mentioned at the beginning of this paper, the fact that no-one can be better off than the child who

⁹⁶ Weinrib (n18) 588.

⁹⁷ *Curlender v Bio-sciences Laboratories* (n12).

⁹⁸ Rosamund Scott, ‘Reconsidering “Wrongful Life” in England after Thirty Years: Legislative Mistakes and Unjustifiable Anomalies’ (2013) 72 *Cambridge Law Journal* 115, 148.

actually exists now means that the calculus of damages is impossible to determine. There is no *better or worse* situation for the child bringing the claim, because ‘nonexistence is not a state that can be given a value’.⁹⁹ This means that determining the condition the child would have been in (and thus being able to determine the damage caused) is not possible. It is clear that the law needs to be changed to end these confusions and contradictions, in particular the injustice of assigning liability to some parties involved in the creation of people (the medical providers) while exempting others from fault (the parents), and for creating liability for a harm that does not occur.

Further problems in England and Wales¹⁰⁰ arise from the CDCLA granting *locus standi* to disabled children to sue the medical provider who caused the disability. Some have claimed that this is discrimination against people with disabilities as it treats disability as harm.¹⁰¹ Now obviously it cannot be discrimination because the foetus that would be aborted is a potential person and so cannot be discriminated against; that is, there is not subject of discrimination. But what the approach mooted here does make clear is that the existence of a disability *cannot* be treated any differently from a ‘normal’ foetus. In other words, disability cannot be conceived of as harming the potential person. Moreover, by adopting these changes, reproductive torts would be much more certain and clearly identify what the harm is, to whom it occurred, and how to remedy or compensate someone for the harm caused. It would also remove the need for the courts to answer questions relating to the value of beings that are brought into existence.

The fundamental problems of comparing existence against non-existence and the impossibility of harming a potential person remain in all cases where a child is allowed to claim. We can now discuss and analyse these reproductive torts in turn, and through this process it will be shown that the considerations that are outlined here apply across jurisdictions, because it is the very nature of allowing children to claim under reproductive torts that gives rise to the problems plaguing the courts and the injustice of creating liability when harm cannot occur.

⁹⁹ Heyd (n15) 30.

¹⁰⁰ This has been corrected from the published version which uses ‘UK’.

¹⁰¹ Lennard Davis, *The Disability Studies Reader* (Routledge 2010).

Wrongful birth

Wrongful birth claims operate on the basis that actual persons, the parents, sought some medical intervention which was carried out incorrectly and thus were denied the option to abort or prevent the conception of a child. The parents therefore seek damages from the medical provider at fault for the harm they have suffered. The only divergence from this definition in English and Welsh law¹⁰² is that the CDCLA allows offspring to claim for a wrongful birth as well as the parents in some circumstances.¹⁰³

In England and Wales¹⁰⁴ the concept of wrongful birth is embodied in case law and legislation, and includes liability for procedures during infertility treatment.¹⁰⁵ Wrongful conception claims in England and Wales¹⁰⁶ are subsumed under the title of wrongful birth cases, thus wrongful conception cases will hereafter be included in the term wrongful birth. However, the CDCLA only deals with situations where something affects the mother or offspring such as physical assault, inaccurate medical advice or a failure to diagnose a disability causing medical condition; basically this covers situations where offspring are ‘born with disabilities which would not otherwise have been present’.¹⁰⁷ Thus the event, in order to qualify under the CDCLA, has to be something that inflicts the disability upon what would otherwise be healthy offspring.¹⁰⁸ Currently it permits offspring to undertake legal action themselves against the medical provider.

One of the most prominent wrongful birth cases (although it would be known as a wrongful conception case elsewhere) is *McFarlane v Tayside Health Board*.¹⁰⁹ This case established that parents may launch a wrongful birth claim when they would not have become pregnant but for negligence on the part of the medical organisation. In *McFarlane* the husband underwent a vasectomy which was unsuccessful and as a result his wife became pregnant. The House of Lords (now the Supreme Court) held that while it would be unjust to have the costs of healthy offspring compensated for, it

¹⁰² This has been corrected from the published version which uses ‘UK law’.

¹⁰³ Congenital Disabilities (Civil Liability) Act 1976 s.1A.

¹⁰⁴ This has been corrected from the published version which uses ‘UK’.

¹⁰⁵ *Ibid* s.1; *McFarlane v Tayside Health Board* (n2); *Whitehouse v Jordan* (n4); *Parkinson v St James and Seacroft University Hospital* [2001] EWCA Civ 530.

¹⁰⁶ This has been corrected from the published version which states ‘UK’.

¹⁰⁷ Congenital Disabilities (Civil Liability) Act s.1(2)(b).

¹⁰⁸ *Ibid* s.1(2).

¹⁰⁹ *McFarlane v Tayside Health Board* (n2).

was fair to have the health authority pay for the damage and distress of an unexpected pregnancy. In a later wrongful birth case, *Parkinson v St James and Seacroft University Hospital NHS Trust*¹¹⁰, the Court of Appeal held that a couple could recover the cost of an unwanted pregnancy due to the failure of a sterilisation procedure and the cost of special care due to the disabilities of their offspring. However the court refused to extend the coverage of *McFarlane* to cover general costs of raising a 'normal' child, therefore restricting compensation to distress and disability. Thus parents can claim for the costs and distress of a pregnancy and the costs of a disability if they are caused by a fault of the medical provider. This is because their interests, rights and goals have been damaged and frustrated by the wrongful action. Although, of course, strictly speaking the normal costs of raising the child should be included, they are excluded on the basis that the birth of a healthy child should not be recoverable because 'society itself must regard the balance as beneficial. It would be repugnant to its own sense of values to do otherwise. It is morally offensive to regard a normal, healthy baby as more trouble and expense than it is worth'.¹¹¹ This rejection is thus based upon the interests of actually existing persons in promoting and valuing the production of children; it is not based on any supposed interests that potential people may have.

This approach has also been taken in some parts of the United States where 'in virtually all cases, courts have awarded the plaintiff mothers ... medical expenses and emotional distress damages' and have rejected damages for the 'cost of raising the unexpected child to adulthood'.¹¹² However unlike in the UK, claims by disabled offspring, like those under the CDCLA, are dealt with differently. In New Jersey case of *Gleitman v Cosgrove*¹¹³ a child was born with defects after the mother contracted rubella, and received incorrect medical advice that rubella did not pose a risk to the foetus.¹¹⁴ When the child was born the parents brought a claim of wrongful birth, matching the situation in England and Wales¹¹⁵, but the child brought a claim of wrongful life not of wrongful birth thus the court permitted a direct claim by the child

¹¹⁰ [2001] EWCA Civ 530.

¹¹¹ *MacFarlane v Tayside Health Board* (n2) Lord Millett 114.

¹¹² Wendy Hensel 'The disabling impact of wrongful birth and wrongful life actions' (2005) 40 *Harvard Civil Rights-Civil Liberties Law Review* 141, 151.

¹¹³ *Gleitman v Cosgrove* (n11).

¹¹⁴ Hensel 155.

¹¹⁵ This has been corrected from the published version which uses 'UK'.

that more closely matches wrongful life cases in England and Wales.¹¹⁶ The case of wrongful life will be considered in the subsequent section, but the divergence should be noted.

There is also the issue regarding the extent to which parents are entitled to information regarding their pregnancy. The relationship between information access about an embryo's genetic condition is discussed by Rosamund Scott in her analysis of the relationship between abortion and wrongful birth.¹¹⁷ That is English and Welsh law¹¹⁸ only serious conditions can justify abortion, so according to Scott in England and Wales¹¹⁹ medical providers only have to convey information when it relates to a serious condition.¹²⁰ For example, if parents would abort on the basis of their embryo having a serious medical condition and the medical provider failed to give them that information that would allow a claim of wrongful birth.¹²¹ If however the information relates to something trivial, say webbed toes, then the parents cannot claim that they would have aborted on that basis and that they should have been informed.

Under the psychological personhood approach suggested here, the embryo would have no standing and so the parents could terminate whenever they wish for whatever reason, even trivial ones. Thus the adoption of psychological personhood would also necessitate a change in abortion law to complete permissibility. Currently the decision of what satisfies the criterion of seriousness is 'based upon the profession's assessment of the incidence of the risk and its seriousness'¹²² but this would no longer be the case because the seriousness of the condition would be irrelevant. There is a strong link between the information provided and the exercising of the right (in the USA) or opportunity (in the UK) to abort; or failing that opportunity suing the medical provider for a wrongful birth. Thus, under the approach proposed here the medical provider would have no justification for withholding information because it is only the interests of the parents that matter; the embryo has no claims to assert against the parents. Thus withholding information that the parents would have aborted on, as the parents are no longer constrained by the 'seriousness'

¹¹⁶ Ibid 155. (This has been corrected from the published version which uses 'UK').

¹¹⁷ Rosamund Scott 'Prenatal screening, autonomy and reasons: the relationship between the law of abortion and wrongful birth' (2003) 11 *Medical Law Review* 265.

¹¹⁸ This has been corrected from the published version which uses 'UK law'.

¹¹⁹ This has been corrected from the published version which uses 'UK'.

¹²⁰ Abortion Act 1967 s.1(1)(d).

¹²¹ Scott, 'Prenatal screening, autonomy and reasons: the relationship between the law of abortion and wrongful birth' 265.

¹²² Ibid 298.

criterion, could lead to a wrongful birth suit. This would greatly increase the scope and power of parental reproductive autonomy and change the law of and relating to abortion.¹²³

As stated earlier, wrongful conception/birth claims are based on the harms are clearly suffered by actual persons to their interests, and would therefore not need to be changed. The basis of wrongful conception/birth claims is coherent and clearly assigns harm and damage both theoretically and legally. However the exception which allows offspring to sue under the CDCLA¹²⁴ must be abolished because it violates the principle that potential persons cannot be harmed by actions which are responsible for constituting their existence. Moreover if the parties involved are in a direct, connected single legal relationship¹²⁵ because the child does not exist at the time the 'harm' is caused. It would also prevent parents from being liable to their offspring, and explain why the current refusal to recognise such liability makes sense because it is only *parental interests* that count. No claims of harm can be recognised in law because persons exist only as a result of those actions. For the most part the law of wrongful conception/birth would remain as it is but on the sole basis of harms to the interests of the parents (actual persons) and without the aberration of the CDCLA.

Wrongful Life

Wrongful life claims are made on the basis that the child who exists does so only because the woman was denied a termination through the fault of the medical provider and that the child's life is so full of suffering that it should not exist. Unlike wrongful conception/birth cases, wrongful life cases are brought by the child themselves not their parents. The essential claim of wrongful life is that the offspring would have been better off not existing. In such cases, the suit for compensation is brought in the name of a (typically) severely disabled child who claims that 'but for' the negligence of the medical provider, they would not have been born at all. This creates a conceptual problem of comparing something with the absence of everything.

¹²³ Abortion will not however be discussed further in this article but this digression demonstrates the wider implications of psychological personhood.

¹²⁴ Congenital Disabilities (n38) s.1A.

¹²⁵ Weinrib (n18) 588.

As Steininger¹²⁶ points out wrongful life claims try to assess the damage inflicted upon offspring from living but do not have an alternative which would be worse as the courts are not prepared to compare existence to non-existence.¹²⁷ Even if the courts were prepared to make such a comparison, it is not clear how such a comparison could be made.

The case of *McKay v Essex Area Health Authority*¹²⁸ made it clear that under English and Welsh law¹²⁹ no offspring can sue for wrongful life. Stephenson LJ stated that it is ‘contrary to public policy, which is to preserve human life, to give a child a right not to be born except as whole, functional being’ and that it is ‘impossible to measure the damages for being born with defects’.¹³⁰ Ackner LJ also says that ‘that there are ... [no damages] in any accepted sense’.¹³¹ More forcefully Griffiths LJ stated that ‘I have come to the firm conclusion that our law cannot recognise a claim for ‘wrongful life’”¹³² for the ‘most compelling reason ... is the intolerable and insoluble problem it would create in the assessment of damages’.¹³³ Thus wrongful life claims have been excluded from English law for two main reasons; firstly that it is against public interest to compare existence with non-existence, and secondly that no damages can be calculated because of the impossibility of comparing loss against not existing. As Ackner LJ stated ‘no comparison is possible and therefore no damage can be established which a court could recognise’.¹³⁴

McKay was decided before the introduction of the CDCLA but even if that had been in force, such a claim would still have been rejected. As Brazier and Cave put it, ‘[w]here the essence of the claim is that the child should never have been born at all, it lies outside the scope of section 1(2)(b)’ of the CDCLA.¹³⁵ More than simply being outside the technical remit of the Act, the Court of Appeal ruled out wrongful life claims because they were against public policy, being a violation of the sanctity of life. Although the sanctity of life is not relevant here, because actions creating

¹²⁶ Barbara Steininger, ‘Wrongful birth and wrongful life: basic questions’ (2010) 1 *Journal of European Tort Law* 125.

¹²⁷ *Ibid* 152.

¹²⁸ *McKay v Essex Area Health Authority* [1982] QB 1166.

¹²⁹ This has been corrected from the published version which uses ‘UK law’.

¹³⁰ *Ibid* 1184.

¹³¹ *Ibid* 1189.

¹³² *Ibid* 1190.

¹³³ *Ibid* 1192

¹³⁴ *Ibid* 1189.

¹³⁵ Margot Brazier and Emma Cave, *Medicine, Patients and the Law* (5th ed, Penguin Books 2011) 321; Congenital Disabilities Act (n39).

potential persons create the life that would be protected by sanctity, accepting life as a potential damage would necessitate ‘an “existence/non-existence” comparison’.¹³⁶ This would be impossible because the court, or indeed anyone, is not capable of determining damages from an assessment which involves comparing existence and non-existence, thus no cause of action was possible.¹³⁷ This view is matched almost exactly in Germany where the Federal Constitutional Court (*Bundesverfassungsgericht*) stated that ‘the child’s existence cannot, in law, be classified as damage’ (*Schwangerschaftsabbruch II* [1993]).¹³⁸

However this strict logical approach has not always been taken. Rosamund Scott, whilst accepting the logical argument regarding existence/non-existence, ‘that there was no-one or no being who could experience non-existence’, argues that this framing of wrongful life claims misses the point.¹³⁹ Scott suggests that the issue of wrongful life claims “lies in *being born under certain conditions*, namely when the burdens of life ... are so severe that they outweigh any compensating goods”¹⁴⁰. This view is similar to the one held by Kirby J who believes that it is the suffering in wrongful life cases that should be given prominence, even going so far as to argue that ‘a life of severe and unremitting suffering is worse than non-existence’.¹⁴¹ In *Curlender v Bio-sciences Laboratories* [1980] Jefferson P.J. claimed that the claimant ‘both exists and suffers, due to the negligence of others’.¹⁴²

In the French case of *Perruche* we can see this suffering-orientated approach at work. The child was allowed to ‘sue his mother’s doctors because they failed to diagnose his condition while she was pregnant, [thus] denying her the choice of an abortion’.¹⁴³ The Cour de Cassation stated that ‘the issue is not his [the offspring’s] birth but his disabilities’.¹⁴⁴ The subsequent public outrage at this decision resulted in the introduction of a new provision in the *code de l’action sociale et des familles* section L114-5 which prohibits offspring bringing a claim where they would not

¹³⁶ Nicolette Priaux ‘Conceptualising harm in the Case of the Unwanted Child’ (2002) 9 *European Journal of Health Law* 337.

¹³⁷ *McKay v Essex Area Health Authority* (n63).

¹³⁸ *Schwangerschaftsabbruch II* [1993] BVerfGE 88, 203.

¹³⁹ Scott, ‘Reconsidering “Wrongful Life” in England after Thirty Years: Legislative Mistakes and Unjustifiable Anomalies’ (n33) 129.

¹⁴⁰ *Ibid* 130.

¹⁴¹ *Harriton v Stephens* (n6) [105].

¹⁴² *Curlender v Bio-sciences Laboratories* (n12) 488.

¹⁴³ Priaux, ‘Conceptualising harm in the Case of the Unwanted Child’ (n72) 339-340.

¹⁴⁴ *Epx Perruche c/ Mutuelle d’assurance du corps sanitaire francais et al* (Cour de Cassation) [2000] JCP 2000 II 10438.

otherwise exist and consequently only parents can now bring a claim under French law.¹⁴⁵ We can thus see that the idea that life itself could be considered ‘damage’ and thus be a cause of action in law was considered unacceptable. Scott’s rejoinder would be that it is the suffering not the life that is compensated for, however this author would still argue that this would fall foul of the ‘actions constituting existence’ problem.

In contrast, in the state of California there was a drift, albeit haphazardly, towards a recognition of a child’s standing to sue. Wendy Hensel¹⁴⁶ provides a clear breakdown of the Californian court’s decisions and reasoning regarding wrongful life. One of the earliest US cases was the previously mentioned case of *Gleitman v Cosgrove* which found that because it was ‘logically impossible’ to ‘measure the difference between life with defects against the utter void of non-existence’ wrongful life claims could not be recognised thus reflecting the position most commonly held in judicial systems.¹⁴⁷ This all changed however in the Californian case of *Curlender v Bio-Science Laboratories* where the Court of Appeal of California allowed offspring to claims for damages relating to ‘damages for the pain and suffering to be endured ... and any special pecuniary loss resulting from the plaintiff’s condition’.¹⁴⁸ Interestingly the court felt that there was ‘no sound public policy reason which should protect those parents from being answerable for the pain, suffering and misery which they have wrought upon their offspring’.¹⁴⁹ Thus in some ways the Californian approach here is the most consistent because it allows ‘harmed’ offspring to claim against all those who brought it into existence; although the basis itself of such a claim is here rejected.

This prompted a swift response from the Californian state legislature to protect parents from such a claim.¹⁵⁰ Shortly thereafter the California Supreme Court in *Turpin v Sortini* confirmed that the sanctity of life did not preclude claims of wrongful life, and allowed damages for ‘significant medical and financial burden’ but rejected claims for general damages relating to the normal costs of existence.¹⁵¹ The claim for

¹⁴⁵ Prialux, ‘Conceptualising harm in the Case of the Unwanted Child’ (n72) 340; *The New York Times*, October 19, 2001, Code de l’action sociale et des familles (amended to include section L114-5 in 2002).

¹⁴⁶ Hensel (n47).

¹⁴⁷ *Ibid* 155.

¹⁴⁸ *Curlender v Bio-sciences Laboratories* (n12) 488.

¹⁴⁹ *Ibid* 487.

¹⁵⁰ The Civil Code of the State of California 1873 s.43.6.

¹⁵¹ *Turpin v Sortini* [1982] 31 Cal. 3d 220, 238.

general damages was rejected ‘because the plaintiff never had a chance to of being born without her affliction, and it would be impossible to ascertain the extent of an injury in this context’.¹⁵² However in most other cases the ‘courts [in the USA] have consistently rejected wrongful life actions’ reasoning that ‘life burdened with defects is better than no life at all and ... that the ... child suffered no legally cognisable injury in being born’ thus ‘damages are incalculable’.¹⁵³ The exceptions (*Curlender* and *Turpin*) in California however ‘have concluded instead that life is not always preferable to non-existence’¹⁵⁴ or that damages are incalculable¹⁵⁵.

Rosamund Scott reaches a similar conclusion, after reviewing the arguments set out by Kirby J in *Harriton v Stephens*, Jefferson P.J. in *Curlender v Bio-sciences Laboratories* and the combined judgement in *Perruche*, regarding how life should be treated in these situations. Scott claims that ‘while it may be helpful to emphasise suffering rather than life *per se* ... care has to be taken that the condition ... be so severe that life will be of sub-zero quality’.¹⁵⁶ Thus Scott’s suggestion is that severity is taken as the basis of wrongful life claims and when a severe enough condition occurs, then a wrongful life claim can be brought against the medical provider. Furthermore, Scott argues that the deployment of the existential threshold in these cases is a legal construction providing a barrier against wrongful life claims. She suggests that ‘emphasising that non-existence cannot be known is ultimately unhelpful, or besides the point ... and loses sight of ... the normative role of the construct of non-existence’.¹⁵⁷ Thus Scott is essentially arguing that justice has lost out to logic.

However Scott’s suggestion is rejected by this author for two reasons. Firstly, the issue of existence-non-existence is crucial because it is this threshold that limits legal actions to those incidents which are cognisable. If we were to accept that non-existence was not a boundary to legal action, then we may be placed in the invidious position of having to take into consideration potential liability owed to people who *may never exist*. Thus the legal prohibition on potential persons from being able to claim, in this case in reproductive torts, should stand. Moreover, the importance of

¹⁵² Hensel (n47)160.

¹⁵³ Ibid 161.

¹⁵⁴ Ibid 162.

¹⁵⁵ Civil Code of California s.43.6.

¹⁵⁶ Scott, ‘Reconsidering “Wrongful Life” in England after Thirty Years: Legislative Mistakes and Unjustifiable Anomalies’ (n33) 133.

¹⁵⁷ Ibid 142.

the logical existential distinction should not be underestimated, as it relates to whether harms have occurred at all and to the existence of a correlative relationship between 'harmer' and 'harmed'.

This leads directly into the second objection to Scott's proposal. If we accept Scott's suggestion, and allow offspring to claim for actions that occur before their existence and use a severity threshold to avoid open-ended exposure to liability just mentioned, this would mean that wrongful life claims are permitting legal action where no harm to the offspring has occurred. It would thus be fundamentally unjust to allow medical providers to be liable for actions which create the particular person who is claiming against them. Medical providers would remain liable to the parents who they failed to treat with the requisite standard of care, but they would not (and should not) be liable to the offspring for creating them. Under the theory of psychological personhood set out here, the action which creates a particular person, that is the act that creates a factual as opposed to merely potential person, cannot be a harmful act to the person created. Thus, cases where wrongful life claims have been recognised should not form the basis of law and should be rejected because they rely on the false premise that life is comparable to non-existence and that those actions which constitute the existence of a person can be classified as beneficial or harmful.

Additionally, the struggle with wrongful life claims, and in some cases recognising them, is down to financial considerations and distributive justice concerns. If there is no liability with the medical provider for disabled children then the costs of their care cannot be taken from the medical provider but must come from another source. The possibility of suing their parents may provide another source of financial security, but practical considerations regarding the payment of damages may prevent claims against parents as they are likely to already be bearing the costs (or some of the costs) but it would be theoretically possible as demonstrated by Californian case of *Curlender v Bio-Sciences Laboratories*. Moreover, as Scott mentions, parental liability would directly conflict with the mother's 'self-determination and bodily integrity' and thus parental liability should be opposed for this reason as well.¹⁵⁸ It thus seems (in the USA at least) that the acceptance of claims was aimed at ensuring some financial provision was in place for the care of the

¹⁵⁸ Ibid 148.

disabled child; this is comparable to the French court allowing the child to sue directly in *Perruche*.

However this paper has concluded that actions which create people cannot be a cause of legal action because being caused to exist cannot be a harm or damage because it is impossible for someone to be harmed by factors constitutive of their existence and identity. Once the child is an actual person, and thus the necessary pre-conditions for them to have interests is met, society (that is the actual persons who make up the social and political entity in which the disabled child lives) do have an obligation to the child, as an actual person. Thus, society must make provision for disabled persons and this can be achieved through general taxation, a welfare system, public healthcare and so on. The child does not have to be able to sue medical providers who cause their existence in order to be financially supported, and indeed it has been argued here that children should not and cannot sue those who cause them to exist.

Conclusion

This article has analysed the basis of reproductive torts wrongful conception/birth and wrongful life through the application of the potential/actual person distinction. The law in England and Wales¹⁵⁹, and in other jurisdictions, on wrongful conception/birth are primarily based upon the interests of the parents who have suffered some loss or harm. This is the only sound basis for these torts. Difficulties and problems arise with wrongful life claims and when wrongful conception/birth claims stray into being claims of the child (as under the CDCLA in England and Wales¹⁶⁰). Claims by the child try to bridge the potential/actual divide and thus allow claims against the medical provider for actions which constitute their very existence. This has been shown to be logically impossible as the person-affecting nature of torts requires that a subject exists who could have been in a better condition. However, when we are discussing actions that cause the existence of a person then the absence of that action results in their *nonexistence*, not their existence in a better state. This argument is clearly articulated when Heyd states that nonexistence is ‘not a state that can be *attributed* to a subject’.¹⁶¹ Consequently the use of personal harm as the basis of

¹⁵⁹ This has been corrected from the published version which uses ‘UK’.

¹⁶⁰ This has been corrected from the published version which uses ‘UK’.

¹⁶¹ Heyd (n15) 30.

reproductive torts has been ruled out as persons cannot be harmed by actions which change who will exist. This means that no potential person has or can claim an interest or a right in either existing or existing in a particular condition.

As the better or worse condition of a subject is a requirement of reproductive torts, a child who results from the actions, however wrongful, cannot be permitted by the law because it contradicts the internal coherence of tort claims. Coherence is important because it allows those subject to the legal system to predict what conduct is permissible and what is not. This knowledge is achieved by the consistent application of the *justification* used to in setting out the law. As Weinrib states the law is ‘not merely a collection of posited norms or an exercise of official power but a social arrangement responsive to moral argument’.¹⁶² Weinrib’s suggests that the ‘[l]aw connects one person to another through the ensemble of concepts, principles, and processes that come into play when a legal claim is asserted’.¹⁶³ However in order for the law to fulfil its function as a set of laws that can successfully be followed, these ‘concepts, principles and processes’ must be internally compatible with each other.¹⁶⁴ As we have seen in the case of reproductive torts, attempts to deal with claims by or the behalf of children have been inconsistent and at times contradictory.

Such inconsistency arises because reproductive torts do not conform to the justification for the existence. That is the crucial relationship requirement whereby ‘the plaintiff and the defendant [are treated] as correlative to one another’ is absent.¹⁶⁵ If reproductive torts are an instrument of ‘corrective justice’ then this relationship is a crucial justification for the existence of reproductive torts in the first place.¹⁶⁶ More importantly as this relationship is the justificatory basis of reproductive tort claims then it should inform the structure of tort claims.

When this relationship is not present then reproductive tort claims cannot be sustained, and in the case of potential persons this requirement is never fulfilled. Thus the structures of reproductive torts are an ad hoc construction rather than a sound coherent and above all justified systematic claim. The changes to reproductive torts set out here are thus a response to this incoherence based upon the consistent

¹⁶² Weinrib (n18) 583.

¹⁶³ Ibid 584.

¹⁶⁴ Ibid 584.

¹⁶⁵ Ibid 593.

¹⁶⁶ Ibid 593.

application of the justificatory theory of persons aimed at forging a ‘harmonious interrelationship among the constituents’ of reproductive torts.¹⁶⁷

The current approach of tort law in judging harm and damage through comparisons between the conditions the person would have been in and the condition they actually are in, is correct because it requires that an actual persons exist who is affected; thus tort law is person-affecting in its approach. However, allowing children to claim under any reproductive torts must logically be prohibited. In the case of England and Wales¹⁶⁸, the current rejection of wrongful life claims must be maintained but the CDCLA must be repealed or reformed to prevent children claiming under the wrongful birth tort. Parents must be able to claim on the basis of negligence or misconduct on the part of medical professionals because they *are* harmed; their interests and rights at the time are affected and harmed. They (as actual persons) could have been in a better situation than the ones they find themselves in. Thus parents and the harm they have suffered can be, and is, the correct basis of reproductive tort claims.

In conclusion this paper advocates a modest legal reform, namely amending the CDCLA to prevent claims by children when actions cause disabilities, but it provides arguments both to engender this reform and to maintain the rejection of any claims by children under reproductive torts (and under wrongful life in particular). The implications of the arguments presented here, both in relation to the law and to the moral and ethical basis of claims, have farther reaching implications in other reproductive issues such as genetic modification, abortion and procreative decisions as in all cases the potential person cannot be the basis of any claims or restrictions on the conduct of the prospective parents. Only the effects on other actual persons or the conduct of the parents themselves (for example on the basis of virtue ethics or a standard of behaviour) can ground legal sanction. Particularly as these different individuals can form correlative relationships with each other that potential and actual persons cannot. These wider implications are beyond the scope this article, and thus would need to be (and deserve to be) considered in separate papers. For the moment this paper concludes with the desirability of reforming the Congenital Disabilities Act suggested earlier and the rejection of any claims by children for actions that are responsible for their existence, because potential persons cannot be affected by

¹⁶⁷ Ibid 593.

¹⁶⁸ This has been corrected from the published version which uses ‘UK’.

actions bringing them to existence. Basically it comes down to this; there is no justification for creating medical provider liability to the offspring for actions which constitute the offspring's very existence.

Part 3

Open questions for a theory of rational constraint

In this thesis I have set out a proposal for an alternative system for constructing legal rules. This alternative system is needed because of the increasing proliferation of conflicting moral systems and the ongoing global integration of nations. These two factors – which can be seen most clearly in developed States dependence on the global economy and their large populations – result in social conditions which have extremely high information costs for interacting with other agents and widespread consequences from our actions that spread beyond those with whom we directly interact. The alternative theory of rational constraint proposed here attempts to provide a responsive method of constructing laws that reflects the goals of those to be regulated so that legal subjects have a self-interested reason for complying with constraints. However, this theory faces some problems when applied to the real world. Here I will outline some of the problems and my present thoughts for resolving these problems which indicate future avenues of research. I will briefly discuss three outstanding problems that need to be resolved in order to present a complete theory of rational constraint:

1. irrational behaviour,
2. multiple optimal outcomes, and
3. the basis for punishing those who might rationally withdraw from the co-operative system

For each of the problems I will first state the problem and then sketch how the problem can be resolved within a theory of rational constraint. I believe that all of the problems are resolvable and that none of them fundamentally undermine the theory of rational constraint.

The most deep rooted problem is that rational constraint theory assumes agents act rationally but it is obvious that human agents are not all that rational. Where does this leave the theory of rational constraint? If agents are not rational it seems rational constraint has no purchase on actual individuals because it ascribes to them something which is not the case. Rational constraint, however, is *normatively correct* for it tells an agent with a set of goals how to behave in order to achieve those

goals. An agent's goals are by definition some state of affairs that they want to bring about thus the prescriptions of practical rationality are things they should follow. The disjunction between the conception of rational agents and the real world occurs because this work is not a sociological study of people rather it is an attempt to provide guidance in solving interaction problems. In most cases the pursuit of these goals will require co-operation with other agents – consequently, agents should (in the normative sense) co-operate. Irrational behaviour, while a fact of life, should be something that agents seek to overcome in order to follow that which best serves the pursuit of their goals – namely, the rules of rational constraint.

Nevertheless, even if we accept that irrationality is a defect that needs to be remedied most agents may not be all that concerned with rationality or with being more rational. But the point of rational constraint is not that it is better per se for people to be rational, rather by using a system of rational constraint to guide their conduct agents can more successfully fulfil their preferences. If it turns out that even with rational constraint guiding action agents are less successful in fulfilling their preferences then this theory fails in its objective. This theory therefore treats rationality as an instrumental practical tool for achieving one's goals in the face of interaction problems. Its normativity derives from the fact that rational constraint is the best means of achieving those outcomes one is already committed to fulfilling – this is what Jonas Olson termed reducible normativity because it is tied to the facts about individual agents (although he does not fully explore this notion).¹ On this basis irrationality needs to be taken into account so that it can be countered by the rules put in place rather than seeking to accommodate those behaviours as part of the system. This indicates that future research will need to focus both on a more realistic model of agents and means of inducing rational over irrational behaviour. We can see that practical rationality has a directive function providing guidance on how to achieve an agent's goals.

The second problem is that practical rationality can generate a number of legal rules and offers a number of operational systems – for example, it may be that a monarchy is just as rational as a democracy or that a law prohibiting prostitution is as rational as a law permitting it. Thus there is a problem with deciding between these different outcomes that distribute costs and benefits differently. As a somewhat rough

¹ Jonas Olson, *Moral Error Theory* (Oxford University Press 2014) 123.

and ready analogy, infertile prospective parents may have a child through either surrogacy or ART – in both cases the couple will end up with a child but the risks and costs are distributed differently. The research I intend to pursue in the future will seek to address this through the use of game theory. My initial thought is that in most circumstances there will clearly be a most rational outcome which gives the greatest fulfilment of preferences for each participating agent. I suspect that this will be aided by the practical limitations on making decision within an already existing social context because the conditions in which a constraint will be imposed will impact its rationality.

But, in principle, with multiple rational outcomes we would have a free choice in which one we adopt as each outcome is the equivalent of the other. For example it does not matter whether we drive on the left or right side of the road as long as everyone drives on the same side. Deciding between equally optimal outcomes requires an additional mechanism to resolve these situations. The options would seem to be either a methodology that always identifies a unique most optimal outcome or some extra-rational basis for choosing between multiple outcomes. I am uncertain regarding extra-rational means of choice – perhaps this will depend on the social system in place (i.e. a democracy might decide through a referendum or the legislature). Other options might to select the outcome that costs the least to achieve from within the current social institutions, the one that imposes the lowest cost on agents or the outcome that confers the largest benefits. In the example of driving, it no longer makes sense for the UK to switch and drive on the right because the left system is already in place. On the other hand it might be more rational to drive on the right because other European countries do and this may minimise accidents amongst tourists (and saving lives may be worth the cost of such a change). In both cases we would need to take into account the entire context in which these decisions are made but only a fully developed theory can resolve this problem because it would be able to account for any contextual scope of analysis. For example, if we introduced the reforms proposed above regarding human rights and reproduction then a total theory would be able to compare the effects of that change on other areas of law and social practice. In other words, a fully theory will make use of the entire social context to select between multiple outcomes that are rational when considered only in a narrower context.

The third significant problem is the fundamental basis of rational constraint which requires conferring a benefit on those constrained. But this suggests that rapists, child abusers and murderers should not comply with the constraints of criminal law if this prevents them from fulfilling their preferences at all. (It may be that the case that rapists, child abusers etc do not have as their greatest preference rape and abuse or it may be that such actions are irrational and thus not preferences at all – but for illustrative purposes let us assume they are their greatest preferences). The key feature of these preferences though is that they cannot be fulfilled with the co-operation of others – these preferences must be forced upon unwilling victims thus there is no mutual benefit in these interactions and victims would devote resources to protect themselves. In the absence of a system of constraint this would lead to violence and conflict. Constraints would be justified on those actions which can only be forced onto unwilling agents because these would protect agents while minimising violence and conflict. Moreover if agents would protect themselves anyway the most efficient means of doing so in a highly anonymous population will be through public enforcement mechanisms. These preferences are thus fundamentally different from preferences that can be fulfilled with the consent of others but which are prohibited. Assisting suicide, voluntary euthanasia, reproduction through surrogates or ARTs – all of these preferences can be fulfilled with the co-operation of others so prohibiting participation in a voluntary action may not be justified at all.

In contrast, preferences that require the non-co-operation of others mean that all potential victims (i.e. the entire population of agents) would want constraints imposed upon those who might act forcibly against them. As constraint is predicated on mutual benefits in any interaction which requires non-co-operation to fulfil a preference will fail to provide a *mutual* benefit. Thus rape should be criminalised because agents will want to be protected from such aggression and criminal sanctions help deter such actions. On the other hand, sado-masochistic sexual activity between consenting adults should not be prohibited because preventing this activity does not protect any other agents as no one is forced to participate. Anyone who refused to abide by this constraints would, in effect, be withdrawing from the co-operative scheme and, consequently, no longer be entitled to its protections. (This risks permitting abuse of criminals but if incarceration is sufficient for eliminating the risk to co-operators it would irrational to take any further measures against them.)

This may apply even in cases where rapists, murders and so on were the majority of the population because the minority (and possibly a significant number of the majority) would not want to be subject to rape, abuse or assault – which requires their non-co-operation – and so it seems that there would always be a reason for agents to set up a criminal system that punishes, deters and reduces incidences of preference fulfilment that requires non-co-operation. While this needs further development the distinction between preferences that can be fulfilled with the co-operation of others and those preferences that can only be fulfilled with the non-co-operation (and perhaps active resistance) of others may play a crucial role in determining the rules of constraint.

Ultimately, the theory of rational constraint needs further research to resolve these situations – a fully developed theory will need, in order to be complete, to solve the problems of irrational behaviour, multiple optimal outcomes and the basis of punishing those who might rationally withdraw from the co-operative system. The purpose of this system remains the minimisation of the enforcement costs of co-operation and the increase of the fruits of collective action while leaving individual agents with as much freedom as possible within the confines of the system of constraints. If rational constraint and its application to agents bring about this kind of outcome then it may be termed a success, if it does not then it will be a failure – either way more work is needed to find out.

Thesis Conclusion

This thesis proposes a non-moral rational choice based framework for justifying legal rules to regulate conduct in populations of anonymous heterogeneous agents. The fundamental question I consider here is why anyone should comply with the rules of a legal system – the answer is because it allows the maintenance of co-operation which confers benefits to all participants although it also requires constraint. As populations increase it becomes harder to have information about those with whom we interact and this permits fragmentation and deviation from historical co-operative systems and their restrictions. For example, it has become increasingly easy to be non-religious in the few decades than it has been in previous centuries (at least in Europe). Another example, from the United States of America concerns demographic changes and the large influx of the Latin American population increasing the diversity of those involved in a national co-operation system. One way of responding to these changes is to increase the amount of constraint imposed, the Saudi Arabian religious police and the repression of dissent in Iran and Russia are good examples of this. Another way is to permit agents as much freedom as possible and to limit state interference - a trend well represented by the introduction of same-sex marriage in the UK and Ireland which grants more freedom to agents while giving them more reason to support the current legal system.

Moreover an oppressive approach is predicated upon the particular cognitivist principles – in Iran and Saudi Arabia these take the form of religious principles, in Russia and its repression of homosexuality these principles are bound up in nationalistic and racial attitudes concerning what Russians are. If we have good reason to doubt the validity of these principles or if they can legitimately be rejected then we face a situation of conflict because individual agents can reject the system imposed upon them. In most cases people seek to bring about change peacefully but not in all (Chechnya is an example of violent resistance to the Russian regime). However, it would be far better for legal systems to reflect the diversity of agents and thus preclude the need for conflict entirely. Accounting for the diversity of agents is extremely difficult and cannot be done within a moral context (because moral systems treat agents as homogeneous). In contemporary populations with mixed and diverse agents resolving interaction problems between themselves without relying on the

oppressive imposition of one particular moral system – which would merely generate resistance and opposition – becomes more important as global integration increases. This is the motivating drive behind this thesis and the effort to develop a new approach to legal systems. Thus the question I hope to have begun to answer is: how can we construct laws so that every agent has a reason to comply with them because they confer benefits upon them that cannot otherwise be achieved and which do not cost agents more than they gain?

This framework is the first step in developing that general strategic theory of law with the aim of resolving the difficulties of interaction between agents who hold competing moral perspectives and preferences yet who need each other to maintain the social milieu in which they live. After considering why moral claims are unable to fulfil their function of justifying constraint – because morality requires prior acceptance of irreducible normativity that can only be asserted – it is proposed that adopting the methodology of rational choice can justify constraint. However, not just any theory of rational choice will do – it must be a theory of rational choice without any foundational moral assumptions to prevent the replication of the problems that apply to moral theories. Consequently, I have attempted to develop a theory that does not rely on moral assumptions by using the descriptive conditions of human existence – that is the limited capacities and vulnerabilities of individuals – as the basis of my system.

Nevertheless it might be that we cannot entirely eliminate moral claims because we have been conditioned by the prevailing moral norms so that moral thinking (of some sort) is embedded in all thinking. Even if this is the case, using the descriptive factors of human existence as our start point can reduce the presence of moral claims as the foundation and methodology for achieving co-operation. To eliminate moral claims we need to use a technical method for identifying optimal rules given facts about those involved in interaction – that is why in future research I will seek to develop a full jurisprudential theory. By using a technical system we can counteract the effects of our intuitions and moral biases because we can model a population of agents as data sets. In order to pursue this non-moral theory, David Gauthier's neo-Hobbesian contractarian framework in his *Morals by Agreement* provides the starting point for my theory of rational constraint as it only contains descriptive assumptions about the behaviour of agents. This is why, for example,

Thomas Scanlon's theory in his '*What We Owe Each Other*'² is excluded because it uses a notion of rationality based upon a prior acceptance of the moral equality of agents.

But, Gauthier's theory is unsatisfactory in one significant respect because it relies solely on an agent's self-interest to maintain co-operative conduct. His theory is one of self-regulation because he treats morality *as voluntary rational compliance* with constraint achieved solely through self-interest – but this precludes the use of second-order constraint which can bring about more efficient outcomes than Gauthier's original theory. As a result of analysing Gauthier's theory I conclude that the best version of a contractarian theory of rational constraint includes second-order constraints based upon non-moral descriptive premises. This thesis attempts to extend the best version of rational choice theory into a legal theory by developing rational choice beyond the traditional economic and, in Gauthier's case, moral (i.e. self-regulation) areas it has been applied to. I term this theory *rational constraint* to distinguish it from other theories of rational choice (particularly Gauthier's) and to indicate the presence of second-order constraints. This is the best version of the theory of rational choice because it explicitly recognises the limits of self-constraint in a large heterogeneous population of highly anonymous agents. Just as importantly the theory of rational constraint places individuals at the heart of the framework by linking constraint to the preferences of individual agents rather than attempting to identify objective rules applicable to all human agents (which inherently treat agents as a homogeneous collective).

The next step is to develop this into a general jurisprudential theory of legal systems and their rules as a means of co-ordinating multiple agents. Thus unlike other uses of rational choice (as in *Game Theory and the Law*³) the purpose of the theory of rational constraint is not confined to an analysis of the efficacy of particular rules used within a pre-determined area of competence. For example, determining whether civil liability in traffic incidents should be fault based or not, whether the costs should primarily be borne by drivers or pedestrians and which set of options reduces the frequency and severity of road accidents fulfils the more limited endeavour of rule-efficiency but fails to justify the presence or existence of civil liability in this area

² (Harvard University Press 2000).

³ Douglas Baird, Robert Gertner and Randal Picker, *Game Theory and the Law* (Harvard University Press 1998).

itself.⁴ Like Richard Posner's economic theory of law rational constraint seeks to justify the existence of law through a rational methodology applied to human behaviour yet unlike Posner's theory rational constraint is not limited to an economic metric – in Posner's case this metric is of wealth-maximisation.⁵ Thus in Posner's theory the purpose of the law is wealth-maximisation while under rational constraint (and following Gauthier) the goal of constraint generally – and also the law – is the maximisation of preference fulfilment.

Constraint is justified when agents need to co-ordinate their activities with each other either to prevent the waste of resources in defence and predation or to produce an outcome with payoffs that – individually – those agents cannot produce. Thus the theory of rational constraint both *analyses* and *justifies* legal systems and rules rather than merely analysing the efficacy of rules or being restricted to an economic theory of interaction. Developing the application of rational choice to produce a general theory of jurisprudence has not been attempted before and this thesis is merely the first step in the production of such a strategic theory of law. The reasons for pursuing this project are threefold – such a theory would explicitly acknowledge and mediate between competing moral beliefs held by heterogeneous agents, it treats the law as a functional process designed promote co-operation (rather than as means to enforce a particular normative moral system) and it would propose a system of law that is dependent on factual features of the human condition rather than on some set of cognitivist moral properties.

One of the main goals of developing this approach is to exclude moral claims from the legal system because basing the law on moral claims contaminates the function of law as a regulatory mechanism by treating all agents as homogeneous moral creatures and giving them a reason to rationally reject constraint. I am not sure if I have been successful in eliminating moral claims entirely at this early stage of developing rational constraint but I do think it is possible. It is important to note that moral claims do inform the construction of legal rules because they are included in an agent's preference but they cannot justify constraint independently of that role. Here, I have sought to begin the process of designing a system of law that does not use any moral claims to justify itself – this contrasts with other legal theories which identify

⁴ See: Ibid 24-28.

⁵ See: Posner R, *Economic Analysis of Law* (Little, Brown & Company 1973) and Posner R, *Economics of Justice* (Harvard University Press 1998).

particular moral claims as components of legal systems and with positivist accounts which do not tell us why we should obey the laws that just happen to exist. Rational constraint theory attempts to eliminate moral contaminants while also providing normative force to legal rules through the methodology of practical rationality and the self-interest of agents whose compliance is necessary for co-operation. Such a legal system would operate as a technical methodology for regulating interaction between agents by identifying the optimal outcome(s) through a strategic analysis given the preferences of agents. As I have not yet explored how effective game theory would be I am unsure whether this would be best method for designing legal rules but as game theory deals with defining outcomes in strategic interaction it seems to be the ideal theory to consider first in further developing this conception of law.

Finally, I then try to apply the basic principles of that theory to specific legal issues surrounding human rights law and reproduction. Through the consideration of human rights law as it relates to reproduction I argue that a more efficient system of human rights would include the power to waive the exercise of the protections that those human rights afford agents. This is a prime example of increasing an agent's capacity to fulfil their preferences without compromising co-operation and the protection of others. Additionally, I argue against the dominant notion of harm as applied to the welfare of future children as a sound basis for justifying the regulation of artificial reproductive technology thus demonstrating the earlier argument that moral claims cannot justify constraint on heterogeneous agents because we have good reasons to be sceptical of moral claims (in this case moral claims concerning future people).

By beginning to develop a non-moral theory of law this thesis proposes an original approach to resolving interaction problems through legal systems. Alternative theories of law rely on moral propositions either in the classical form of natural law theories, through reliance on a particular moral conception (for example, an assumed basic moral equality between all agents as in Scanlon) or an economic conception of law as in Richard Posner's economic theory of law.⁶ The rational constraint theory of law does not seek to identify the *correct* or *right* set of immutable laws nor does it seek to identify a single absolute authoritative source of law (as John Austin's

⁶ See: Posner R, *Economic Analysis of Law* (Little, Brown & Company 1973) and Posner R, *Economics of Justice* (Harvard University Press 1998).

command theory sought to).⁷ Even the notion of compliance does not authorise law so much as give concrete enforcement to a particular set of rules through the decisions of individual agents in response to that very system of rules.

The development of a non-moral legal system is urgently needed in the contemporary globally integrated world with increasingly heterogeneous populations. As populations become more heterogeneous they become more fragmented and the dominance of any one moral system diminishes. The rational constraint theory of law is aimed at minimising the conflict that occurs amongst such a population by crafting legal rules that permit the most freedom to individuals without compromising co-operation – in other words by crafting those legal rules that are necessary for social cohesion and for producing optimal outcomes. Responding to the breakdown of dominant moral narratives is the reason for using the system set out here. As rational constraint is context dependent it can adapt to changing circumstances but this adaptation is guided by the concept of achieving the most rationally optimal outcome – that is the most efficient system in terms of consumption of resources in enforcement and preference fulfilment.

The theory of rational constraint is, therefore, a new and radical approach to constructing legal systems – it is new and radical because it prioritises the adaptation of law to the population it governs, it uses the preferences of individuals to determine legal rules and it uses non-moral rational choice rather than one or more moral premises to justify the existence of constraints on an agent's conduct. If this can be developed into a fully fledged legal theory it would provide a method for determining legal rules directly and, possibly, independently of political organs such as legislatures and executive administrations (with the possible exception of choosing between multiple equally rational outcomes). Consequently, the theory of rational constraint proposes a technical methodology for constructing legal rules to regulate interaction between agents producing a legal system for the increasingly heterogeneous complex and integrated world in which we live.

⁷ Austin J, *'The Province of Jurisprudence Determined'* (first published 1832, Hackett Publishing Company 1998).

Bibliography

Articles and Books

- Ainslie G, *Breakdown of Will* (Cambridge University Press 2001)
- Austin J, *The Province of Jurisprudence Determined* (first published 1832, Hackett Publishing Company 1998)
- Ayer A, *Language, Truth and Logic* (2nd edn, Gollancz 1946)
- Baier K, *The Moral Point of View* (Random House 1966)
- Baier K, *The Rational and the Moral Order* (Open Court 1995)
- Baird D, Gertner R and Picker R, *Game Theory and the Law* (Harvard University Press 1998)
- Baker L, *Persons and Bodies* (Cambridge University Press 2000)
- Balkin J, 'Too Good To Be True: The Positive Economic Theory of Law' (1987) *Faculty Scholarship Series Paper 290*
- Benatar D, 'Why it is Better Never to Come into Existence' (1997) 34 *American Philosophical Quarterly* 345
- Bennett R, 'The Fallacy of the Principle of Procreative Beneficence' (2009) 23 *Bioethics* 265
- Berlin I, *The Proper Study of Mankind: An Anthology of Essays* (Random House Publishing 1998)
- Blackburn S, *Spreading the Word* (Oxford University Press 1984)
- David Boonin, *The Non-Identity Problem and the Ethics of Future People* (Oxford University Press 2014)
- Braithwaite R, *Theory of Games as a tool for the Moral Philosopher* (Cambridge University Press 1955)
- Brassington I, *Truth and Normativity* (Ashgate Publishing Company 2007)
- Brazier M and Cave E, *Medicine, Patients and the Law* (5th ed, Penguin Books 2011)
- Broome J, 'Is Rationality Normative?' (2007) 2 *Disputatio* Special Issue 161
- Buchanan A, Brock D, Daniels N and Wikler D, *From Chance to Choice* (Cambridge University Press 2000)
- Buchanan A, *The Heart of Human Rights* (Oxford University Press, 2014)
- Craig P and de Buca G, *EU Law* (5th edn, Oxford University Press 2011)
- David Copp, 'Contractarianism and Moral Skepticism' in Peter Vallentyne (ed) *Contractarianism and Rational Choice* (Cambridge University Press 1991)

Danielson P, 'Closing the Compliance Dilemma' in Peter Vallentyne (ed) *Contractarianism and Rational Choice* (Cambridge University Press 1991)

Davis L, *The Disability Studies Reader* (Routledge 2010)

Dawkins R, *The Selfish Gene* (Oxford University Press 2006)

Donnelly J, *Universal Human Rights in Theory and Practice* (Cornell University Press 2003)

Dworkin R, *Taking Rights Seriously* (Bloomsbury Academic 2013)

Ebrahim A, *Biomedical Issues: Islamic Perspective* (A.S. Noordeen 1993)

Editorial 'The Guardian view on criminalising drinking during pregnancy: no cheers' *The Guardian* 5.11.2014

Eijkholt M, 'The Right to Found a Family as a Stillborn Right to Procreate?', 18 *Medical Law Review* (2010) 127

Elliott C and Quinn F, *Tort Law* (Longman, 2011)

Elster J, *Ulysses and the Sirens* (Cambridge University Press 1984)

Elster J, *Ulysses Unbound* (Cambridge University Press 2000)

Eriksson M.K., *The Right to Marry and to Found a Family: A World-Wide Human Right* (Coronet Books, Inc, 1990)

Feinberg J, 'Voluntary Euthanasia and the Inalienable Right to Life (1978) 7 *Philosophy and Public Affairs* 93

Feinberg J, *Harm to Others* (Oxford University Press 1987)

Fisher A and Kirchin S (eds), *Arguing About Metaethics* (Routledge 2006)

Fletcher J, 'Indicators of Humanhood: a tentative profile of man' (1972) 2(5) *The Hastings Center Report* 1

Fuller L, *The Morality of Law* (Yale University Press, 1964)

Gauthier D, 'Why Contractarianism?' in Peter Vallentyne (ed) *Contractarianism and Rational Choice* (Cambridge University Press 1991)

Gauthier D, *Morals by Agreement* (Clarendon Press 2001)

Gauthier D, *The Logic of Leviathan* (Clarendon Press 2011)

Giubilini A and Minerva F, 'After-birth abortion: Why the baby should live?' (2012) 39 *Journal of Medical Ethics*

Glannon W, 'Genes, Embryos and Future People' (2002) 12 *Bioethics* 187

Gledhill R, 'Church attendance has been propped up by immigrants, says study' (2014) *The Guardian* 3 June

Glover J, *Choosing children: Genes, disability and design* (Oxford University Press 2008)

Habermas J, 'The Theory of Communicative Action: Reason and the Rationalisation of Society v.1' (Polity Press 2004)

Habermas J, 'The Theory of Communicative Action: Lifeworld and Systems, a Critique of Functionalist Reason v.2' (Polity Press 2004)

Hampton J, 'Two Faces of Contractarian Thought' in Peter Vallentyne (ed), *Contractarianism and Rational Choice* (Cambridge University Press 1991)

Hanser M, 'The Metaphysics of Harm' (2008) 77 *Philosophy and Phenomenological Research* 421

Harman G, *The Nature of Morality* (Oxford University Press 1977)

Harman G, 'Moral explanations of natural facts – can moral claims be tested against reality?' (1986) 24 *Southern Journal of Philosophy Supplement*

Harman E, 'Can We Harm and Benefit in Creating?' (2004) 18 *Philosophical Perspectives* 89

Harris J, *Wonderwoman and Superman* (Oxford University Press 1993)

Harris J, *Enhancing Evolution* (Princeton University Press 2007)

Hart H, *The Concept of Law* (Oxford University Press, 2012)

Hensel W, 'The disabling impact of wrongful birth and wrongful life actions' (2005) 40 *Harvard Civil Rights-Civil Liberties Law Review* 141

Heyd D, *Genethics: Moral Issues in the Creation of People* (University of California Press 1992)

Hobbes T, *Leviathan* (first published 1651, Oxford University Press 1996)

Hohfeld W, 'Some Fundamental Legal Conceptions as Applied to Judicial Reasoning' (1913) 23 *The Yale Law Journal* 16

Hoppe N and Miola J, *Medical Law and Medical Ethics* (Cambridge University Press 2014)

Hume D, *Enquires Concerning Human Understanding and Concerning the Principles of Morals* (first published 1751, Oxford University Press 1975)

Ignateiff M, *Human Rights as Politics and Idolatry* (Princeton University Press 2011)

Jackson E, *Regulating Reproduction* (Hart Publishing 2001)

Jackson E, 'Conception and the Irrelevance of the Welfare Principle' (2002) *Modern Law Review* 65:2

Kass L, 'Ageless Bodies, Happy Souls' [2003] *The New Atlantic* 9)

Kavka G, 'The Paradox of Future Individuals' (1982) 11 *Philosophy and Public Affairs* 93, 93

Kirchin S, *Metaethics* (Palgrave Macmillan 2012)

Kraus J and Coleman J, 'Morality and the Theory of Rational Choice' in Peter Vallentyne (ed) *Contractarianism and Rational Choice* (Cambridge University Press 1991)

Lee E, Macvarish J & Sheldon S, 'Assessing Child Welfare Under the Human Fertilisation and Embryology Act 2008: A Case Study in Medicalisation?' (2014) 36(4) *Sociology of Health and Illness* 500

Leeson P, *Anarchy Unbound* (Cambridge University Press 2014)

Locke J, *An Essay Concerning Human Understanding* (first published 1689, Everyman 1996)

Lomasky L, *Persons, Rights and the Moral Community* (Oxford University Press 1990)

Luce R and Raiffa H, *Games and Decisions* (John Wiley and Sons 1957)

Mackie J, *Ethics: Inventing Right and Wrong* (Penguin Books 1990)

Mason J and Laurie G, *Law and Medical Ethics* (Oxford University Press 2006)

Sheila McLean, 'The Right to Reproduce' in Thomas Campbell and others, eds., *Human Rights: From Rhetoric to Reality* (Basil Blackwell, 1986)

Mill J, *On Liberty* (Cosimo Classics 2005)

Miller A, *Contemporary Metaethics* (Polity Press 2013)

Morris C, 'Moral Standing and Rational-Choice Contractarianism' Projects' in Peter Vallentyne (ed) *Contractarianism and Rational Choice* (Cambridge University Press 1991)

Narveson J, 'Contractarian Rights' in R. G. Frey (ed) *Utility and Rights* 1984 (University of Minnesota 1984)

Olson J, *Moral Error Theory* (Oxford University Press 2014)

O'Neill O, 'The Dark Side of Human Rights' (2005) 81 *International Affairs* 427

Parfit D, *Reasons and Persons* 1984 (Clarendon Press 1984)

Parfit D, *On What Matters*, vol 2 (Oxford University Press 2011)

Posner R, *Economic Analysis of Law* (Little, Brown & Company 1973)

Posner R, *Economics of Justice* (Harvard University Press 1998)

Priaulx N, 'Conceptualising harm in the Case of the Unwanted Child' (2002) 9 *European Journal of Health Law* 337

Priaulx N, 'Health, Disability & Parental Interests: Adopting a Contextual Approach in the Reproductive Torts' (2005) 12 *European Journal of Health Law* 213

Rawls J, *A Theory of Justice* (rev edn, Harvard University Press 1999)

Railton P, 'Moral Realism' (1986) 95 *Philosophical Review* 163

Richards B, 'Inalienable Rights: Recent Criticism and Old Doctrine' (1969) 29 *Philosophy and Phenomenological Research* 391

Robertson J, *Children of Choice* (Princeton University Press 1996)

Rohwedder D, 'Ferguson v. McKiernan: Can a Sperm Donor Be Held Liable for Child Support After the Recipient Has Contractually Waived that Right?' (2008) 32 *American Journal of Trial Advocacy* 229

Rousseau J, *The Social Contract* (first published 1762, Oxford University Press 2008)

Sayre-McCord G, 'Deception and Reasons to be Moral' in Peter Vallentyne (ed) *Contractarianism and Rational Choice* (Cambridge University Press 1991)

Savulescu J, 'Procreative Beneficence: why we should select the best children' (2001) 15 *Bioethics* 413

Savulescu J, 'Procreative Beneficence: reason to no have disabled children' in Loane Skene and Janna Thompson (eds) *The Sorting Society* (Cambridge University Press 2008) 51

Savulescu J and Kahane G, 'The Moral Obligation to Create Children with the Best Chance of the Best Life' (2009) 23 *Bioethics* 274

Scanlon T, *What We Owe Each Other* (Harvard University Press 2000)

Scott R, 'Prenatal screening, autonomy and reasons: the relationship between the law of abortion and wrongful birth' (2003) 11 *Medical Law Review* 265

Scott R, 'Reconsidering "Wrongful Life" in England after Thirty Years: Legislative Mistakes and Unjustifiable Anomalies' (2013) 72 *Cambridge Law Journal* 115

Shoemaker S and Swinburne R, *Personal Identity* (Blackwell 1985)

Sidgwick H, *Methods of Ethics* (7th edn, University of Chicago Press 1907)

Sieghart P, *The International Law of Human Rights* (Clarendon Press, 1983)

Singer P, *Practical Ethics* (3rd edn, Cambridge University Press 2011)

Slors M, *The Diachronic Mind* (Kluwer Academic Publishers 2001)

Smith H, 'Deriving Morality from Rationality' in Peter Vallentyne (ed) *Contractarianism and Rational Choice* (Cambridge University Press 1991)

Smith M, *The Moral Problem* (Blackwell Publishing 1994)

Sobel D, 'Full Information Accounts of Well-being' (1994) 104 *Ethics* 784

Steinbock B, 'Wrongful Life and Procreative Decisions' in Melinda Roberts & David Wasserman (eds), *Harming Future Persons: Ethics, Genetics and the Nonidentity Problem* (Springer 2009)

Steinbock B, *Life Before Birth* (2nd edn, Oxford University Press 2011)

Steininger B, 'Wrongful birth and wrongful life: basic questions' (2010) 1 *Journal of European Tort Law* 125

Sturgeon N, 'Harman on moral explanations of natural facts' (1986) 24 *Southern Journal of Philosophy Supplement* 69

Sumner L, *The Moral Foundation of Rights* (Clarendon Press 1987)

Tooley M, 'Abortion and Infanticide' (1972) 2 *Philosophy and Public Affairs* 37

Wacks R, *Understanding Jurisprudence* (Oxford University Press 2009)

Weinrib R, 'The Jurisprudence of Legal Formalism' (1993) 16 *Harvard Journal of Law and Public Policy* 583

Wellman C, *A Theory of Rights* (Rowman & Allanheld 1985)

Wilkinson S, *Choosing Tomorrow's Children* (Oxford University Press 2010)

Wright C, 'Moral values, projection, and secondary qualities' (1988) 62 *Proceedings of the Aristotelian Society Supplement* 11

Vallentyne P, 'Gauthier's Three Projects' in Peter Vallentyne (ed), *Contractarianism and Rational Choice* (Cambridge University Press 1991)

Veer D, 'Are Human Rights Alienable?' (1980) 37 *Philosophical Studies* 165

Unknown, 'Code de l'action sociale et des familles (amended to include section L114-5 in 2002)' (2001) *The New York Times* 19 October

Legislation

Abortion Act 1967

Basic Law of the Federal Republic of Germany

Civil Code of the State of California 1873

Code de l'action sociale et des familles (amended to include section L114-5 in 2002)

Congenital Disabilities (Civil Liability) Act 1976

Human Fertilisation and Embryology Act 1990

Human Fertilisation and Embryology Act 2008

Surrogacy Arrangements Act 1985

State and Regulatory documents

Human Fertilisation and Embryology Authority, '8th Code of Practice' (2012)

Office of National Statistics, 2001 Census

Office of National Statistics, 2011 Census

Commission on Assisted Dying, 'Final Report' (2012)

Cases

Attorney-General's Reference (No.3 of 1994) [1998] AC 245

Criminal Injuries Compensation Authority v First-tier Tribunal [2014] EWCA Civ 1554

Curlender v Biosciences Laboratories [1980] 165 Cal. Rptr. 477

Dickson v the United Kingdom [2008] 46 EHRR 41

Epx Perruche c/ Mutuelle d'assurance du corps sanitaire francais et al (Cour de Cassation) [2000] JCP 2000 II 10438

Evans v Amicus Healthcare Ltd [2004] 2 W.L.R. 713

Evans v the United Kingdom (2006) 43 E.H.R.R. 21

Farrar v King's Healthcare NHS Trust [2010] 1 W.L.R. 2139

Ferguson v McKiernan 855 A.2d 121 (Pa. Super 2004)

Gleitman v Cosgrove [1967] 49 N.J. 22

Harriton v Stephens [2006] Unreported Cases High Court of Australia

Hass v Switzerland (2011) 53 EHRR 1169

JP v LP, SP, CP [2014] EWHC 595 (Fam)

MacFarlane v Tayside Healthboard [2000] 2 AC 59

McKay v Essex Area Health Authority [1982] Q.B. 1166

Parkinson v St James and Seacroft University Hospital [2001] EWCA Civ 530

The Queen on the Application of Dianne Pretty v Director of Public Prosecutions [2001] WL 1171775 (Queen's Bench)

R. (on the application of Pretty) v Director of Public Prosecutions [2001] UKHL 61

R. (on application of Purdy) v Director of Public Prosecutions [2009] UKHL 45

R. (on the application of Nicklinson) v Ministry of Justice [2013] EWCA Civ 961 (Appeal Court)

Regina (Nicklinson) and another v Ministry of Justice [2014] UKSC 38

Re P (Minors) (Wardship: Surrogacy) [1987] 2 F.L.R. 421 Case Analysis

Re X (A child) (Surrogacy: Time Limit) [2014] EWHC 3135 (Fam)

Schwangerschaftsabbruch II [1993] BVerfGE 88

Turpin v Sortini [1982] 31 Cal. 3d 220

Vo v France (2005) 40 E.H.R.R. 12

W v H (child abduction: surrogacy)(No1) [2002] 1 F.L.R. 1008

Whitehouse v Jordan [1981] W.L.R. 246

Media

Dignitas accessed 14/12/2014

http://www.dignitas.ch/index.php?option=com_content&view=article&id=22&lang=en

Human Fertilisation and Embryology Authority website (accessed 27/12/2015)

<http://www.hfea.gov.uk/2588.html>

Leone S, *Once upon a time in America* 1984

Voluntary Human Extinction Movement <http://www.vhemt.org/>

Appendix

The following appendix contains the published original version of chapter four as ‘Potential persons and the welfare of the (potential) child test’ (2014) 14 *Medical Law International* and the published chapter five Samuel Walker ‘Applying the actual/potential person distinction to reproductive torts’ (2014) *Medical Law International*.

Potential persons and the welfare of the (potential) child test

Medical Law International
2014, Vol. 14(3) 157–171
© The Author(s) 2014
Reprints and permission:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/0968533214550117
mli.sagepub.com



Samuel E. P. Walker

University of Manchester, UK

Abstract

This article considers and critiques the theoretical basis of section 13(5) of the Human Fertilisation and Embryology Act 1990 (HFE Act),¹ which requires the welfare of the potential child to be taken into account when considering whether parents should be given access to fertility treatment. It will be argued that potential persons, that is, persons who do not yet exist, have no claims, interests or standing that can restrict the actions of actual persons. This claim will be based upon the necessity of existence before things can be said to *affect a person*. As persons are the subjects for whom ‘good’ and ‘bad’ outcomes apply, actions which establish the preconditions for their existence cannot be subjected to considerations of the effect on the potential person. This is because potential persons are not *affected* by actions, but are the *consequence* of actions. Prospective parents, for example, should not be prohibited from having disabled offspring on the basis of the *effect on the child*, as different decisions relating to that child will change which persons exist, and thus the necessary preconditions for value will change. Based upon this existential framework, it will be argued that only the interests of *actual* persons can constrain the actions of those involved in reproduction. Thus, the HFE Act’s formulation of the welfare test must be repealed and only the interests of prospective parents and other actually existing people should be capable of constraining reproduction.

Keywords

Genetic, person, embryo, existence, welfare test

1. As amended by the 2008 Act of the same name.

Corresponding author:

Samuel E. P. Walker, School of Law, Faculty of Humanities, University of Manchester, Williamson Building, Oxford Road, Manchester M13 9PL, UK.

Email: sep.walker@gmail.com

Introduction

Developments in fertility treatment and artificial reproductive technology have created an unprecedented level of control over the process of reproduction. Prospective parents now have many options for fulfilling their desire to have children; from artificial insemination to *in vitro* fertilization, from adoption to surrogacy and possibly through genetic modification. In response to the development of these technologies, the UK government established the Committee of Inquiry into Human Fertilisation and Embryology (1982–1984) to consider the implications and responses to these emerging reproductive technologies. The Committee produced a report (also known as the Warnock Report²), which then formed the basis for a draft bill that was passed into law as the Human Fertilisation and Embryology Act 1990 (hereafter the HFE Act). This Act created the Human Fertilisation and Embryology Authority, which has regulatory oversight of all fertility treatment and reproductive services (although this regulatory function may be transferred to the Care Quality Commission and others).³ The HFE Act contains a provision in section 13(5), which requires that ‘the welfare of any child who may be born as a result of the treatment’ be taken into account.⁴ It is this ‘welfare of the child test’ as it exists in the HFE Act that is questioned and critiqued in this article.

It will be argued that the welfare test when applied to the *potential* child is problematic because it is too burdensome on prospective parents (hereafter parents) both because it requires that the parents are vetted to determine whether they are ‘suitable’ and because the test imposes obligations on the parents for the sake of the child (most noticeably in the case of disability). Moreover, if these obligations are complied with it may require the parents to abort the foetus or otherwise not carry through the pregnancy, thus the child will never come to exist; and this would require that parents act for a child that will never exist. Equally, if genetic modification technology became available, it would require that parents use it to improve the welfare outcome of the potential child. The welfare of the potential child test therefore creates a discriminatory system in which those who cannot reproduce naturally are doubly penalized; they must prove their competence to be parents, and the test creates a strong prejudice against parents who want to produce a child who has, or might have, a disability. These restrictions thus need to be firmly grounded in order to justify the imposition of onerous obligations on parents and the denial of services to those who may not wish to undergo genetic testing for disability or those who may want to have a child with a particular disability.

The justification for restricting access to reproductive technology, as shown by section 13(5), is the ‘welfare of any [potential] child who may be born’.⁵ It is the benefits and welfare that accrue to the potential child that justify considering the conduct of the parents and the risk of disability. However, it will be argued that consideration of the welfare of the potential child is not valid because it is incorrect that a child can be benefitted or harmed

2. Department of Health and Social Security, ‘Report of the Committee of Inquiry into Human Fertilisation and Embryology’, 1984.
3. Human Fertilisation and Embryology Authority, ‘Statement: HFEA Responds to Government’s Proposals on the Future of the HFEA’, 2012.
4. HFE Act 1990, section 13(5).
5. HFE Act, section 13(5).

by actions that cause them to exist; that is, actions which *constitute the person cannot affect that same person* because their existence is a *necessary* precondition for them to benefit or be harmed. As David Heyd states, ‘they [potential persons] do not meet certain preconditions of existence and identity’,⁶ which would allow them to be the object of beneficence or maleficence. Thus the welfare of the potential child cannot be the basis of claims regarding benefitting or harming the child; therefore, the considerations that are used to restrict access to reproductive technology and service are unfounded. This becomes even more apparent when the denial of access to reproductive technology services means that no child *ever exists*. As the child never exists, it is logically impossible for them to be ‘benefitted’ or ‘harmed’, but this also applies in reverse, that is, a child who does come to exist cannot be benefitted or harmed before their existence either. This is what Heyd calls ‘full symmetry’ between benefitting and harming in both cases of not bringing and of bringing a child into existence.

Consequently, the law relating to reproductive technology and services ought to be reformed by removing the ‘welfare of the potential child’ test from the HFE Act, as it is based upon a concept of harm in which harm is impossible. Using the example of genetic technology⁷ available to parents (such a genetic testing, screening, selection etc.⁸), the implications of this test for what is permissible will be explored, and the scope of parents’ decisions will be explicated. These implications do not however suggest that parents should have unrestricted access to these services (because resource constraints and distributive justice are also involved), but they do mean that parents cannot be denied access to genetic modification services on the basis of harming a child who will never exist and who cannot be harmed if they do exist. Only actual persons can be affected by the actions of the parents; for example, through increasing burdens on health care services by allowing parents to have disabled offspring. Furthermore, the welfare of the child once it actually exists is covered by the ‘welfare of the child test’ found in section 1 of the Children Act 1989, thus providing a means for dealing with parents if they are abusive to their offspring. Society could therefore decide that people who want to be parents have to conform to a certain standard of conduct and comport themselves in a certain

-
6. D. Heyd, *Genethics: Moral Issues in the Creation of People* (Oxford: University of California Press, 1st ed., 1992), p. 36.
 7. Genetic technology here encompasses all the actions parents might take to ensure that their future offspring either possesses or does not possess a particular genetic trait: see ‘New Down’s syndrome blood test more reliable, say researchers’ *The Guardian*, 7 June 2013; ‘Lesbian couple have deaf baby by choice’ *The Guardian*, 8 April 2002; ‘IVF baby born using revolutionary genetic-screening process’ *The Guardian*, 7 July 2013; ‘Families hope “Frankenstein science” lobby will not stop gene cure for mitochondrial disease’ *The Guardian*, 15 February 2014.
 8. For more work on the issues of genetic technology, see A. Buchanan et al. *From Chance to Choice* (Cambridge: Cambridge University Press, 2001); J. Glover, *Choosing Children: Genes, Disability and Design* (Oxford: Oxford University Press, 2008); J. Harris, *Enhancing Evolution: The Ethical Case for Making People Better* (Princeton, Princeton University Press, 2007); L. Skene (ed.), *The Sorting Society: The Ethics of Genetic Screening and Therapy* (Cambridge: Cambridge University Press, 2008).

manner, but this would have to apply to *all* parents (whether using reproductive technology or conceiving naturally) because the determination and judgement is located in the parents themselves.⁹ Currently, those seeking medical assistance to have children are subject, in effect, to a competency type test which is not applied to the rest of the population. The arguments that will now be presented will support the final recommendations of this article, namely that the welfare of the potential child test in the HFE Act must be removed and that the determination of access to reproductive services should be based upon the effect on actual persons or a standard applicable to all.

The welfare of the (potential) child test

Under section 13(5) of the HFE Act, the welfare of the child test operates as follows:

A woman shall not be provided with services . . . unless account has been taken of the welfare of the child who may be born as a result of the treatment (including the need of that child for supportive parenting), and of any other child who may be affected by the birth

The HFE Authority's 8th Code of Practice guidance notes on the welfare of the child state that the 'factors to be taken into account during the assessment process' of the parents include the risk of 'serious physical or psychological harm or neglect'¹⁰ and 'where the medical history indicates that any child who may be born is likely to suffer from a serious medical condition'.¹¹ Subsections 13(9) and 13(10) of the Act, inserted by the amending 2008 Act, state that 'embryos known to have' a serious disability, illness or 'other serious medical condition' 'must not be preferred to those [embryos] that are not known to have such an abnormality'.¹² Together, these legislative provisions and code of practice notes make two things clear (1) that the conduct of parents is subject to scrutiny and (2) that parents cannot select *for* an embryo that has a disability but can reject an embryo on that basis. As both sets of considerations operate under the welfare of the (potential) child test, it is clear that the reasons and justifications for these considerations are rooted in assumptions that they are *harmful to the child's welfare*.

Thus a positive determination that the potential child will be subject to abuse or neglect, or will have a serious disability, will prevent the parents from accessing reproductive services, frustrating in its entirety their interest and desire to have children. This places parents in the invidious position of not being able to have children without medical help but also of being judged both on their characteristics and their choice of foetus. Moreover, if any of these judgements go against them, then the child they would

9. As far as I am aware, no one has argued that all parents should prove their competency to raise children. There are a number of programs, such as the Triple-P program (Positive Parenting Program), which are designed to give parents strategies for raising children, although these are obviously voluntary.

10. Human Fertilisation and Embryology Authority, Code of Practice, 8th ed., 2009, section 8.10, Available at: <http://www.hfea.gov.uk/3401.html> (accessed 21 July 2014).

11. Code of Practice, section 8.10(b)(iii).

12. Human Fertilisation and Embryology Act 2008, section 14.

supposedly be harming would never even exist. Persons who do not exist cannot be harmed because their non-existence precludes anything affecting them in any sense. We will come to the impossibility of harming potential persons in due course. For now, we can note the effects of the current legal regime. Firstly, by requiring those who seek reproductive services to prove their ability to be 'good' parents, those who are already at a severe disadvantage are further penalized. The institutional burden imposed on them, which is not imposed on other members of society, could be construed as an unacceptable interference with the parent's reproductive autonomy. Some have argued that a right to reproductive autonomy can be found under Articles 8 and 12 of the European Convention of Human Rights, and interference with this right would need to be strongly justified. The discussion of reproductive autonomy, and whether there is a right to it, is beyond the scope of this article.¹³ However, it should be noted that the arguments presented here mean that the requirement of proving parental capability would not make sense *on the basis of the potential child's welfare*. Requiring proof of parenting competence could be justified if applied to all potential parents, but it would have to be a universal requirement based upon the interests of actual members of society. Thus the justification for the burden of proving parental capability would be focused solely on the parents and would not be related to the supposed characteristics of the potential child.

More importantly though, the demand that disabled embryos are not to be preferred and the fact that disability counts against the potential child's welfare implies that there is an obligation *not* to have disabled children. This obligation can be inferred because the logical corollary of the statement that disabled foetuses 'are not be preferred'¹⁴ is that non-disabled foetuses *are to be preferred*. This restriction on having disabled offspring removes the choice of parents and mandates what they can and cannot do. Furthermore, if genetic modification could improve the potential child's 'welfare', then it may be required of parents to use genetic technology because the basis of the constraints are the potential child's interests which (as shown by the restrictions imposed on parents by the HFE Act now) are strong enough to force parents to refrain from acting (and thus logically can require them to act) in certain ways.

As the possibility of requiring parents to use genetic modification of the embryo would be justified on the basis that disability is against the potential child's welfare then, if this justification is incoherent, the preference for non-disabled offspring and the constraint imposed on the parents must be removed. Parents could therefore not be prevented from having disabled offspring if they so wished, at least on the basis that it is against the child's interests. The welfare of the potential child then is the source of the problems that

13. For the legal consideration of a right to reproduce, see *Dickson v. UK* [2008] 46 EHRR 41; *Evans v. UK* [2008] 46 EHRR 34. For further discussion, see M. Eijkholt, 'The Right to Found a Family as a Stillborn Right to Procreate?', *Medical Law Review* 18 (2010), p. 127; M.K. Eriksson, *The Right to Marry and to Found a Family: A World-Wide Human Right* (Philadelphia: Coronet Books, Inc, 1990); E. Jackson, *Regulating Reproduction: Law, Technology and Autonomy* (Oxford: Hart, 2001); S. McLean, 'The Right to Reproduce', in T. Campbell and others, eds., *Human Rights: From Rhetoric to Reality* (Oxford, Basil Blackwell, 1986), p. 103.

14. Human Fertilisation and Embryology Act 2008, section 14.

are thrown up by applying the welfare test to those seeking access to reproductive technology; namely that the potential child welfare acts as justification for restricting who can have children and under what conditions, when in fact the child's welfare *cannot be affected* by any decision made. It is to this person-affecting dilemma to which we now turn.

Welfare of the child: What child?

The fundamental problem with the welfare of the (potential) child test is that welfare by definition is *for someone*, but in this case the person concerned does not exist. The welfare of the child test, contained in the HFE Act section 13(5), explicitly includes the concept of harm to potential persons, thus identifying the potential child as the *subject affected*. This is, however, a logical impossibility because the necessary condition for welfare to be considered, the existence of a person, is not met. Moreover, if the potential person's welfare is considered to be at risk, then they will never come to exist so they will never benefit from this welfare decision. As welfare is 'person affecting', it is predicated upon the existence of someone for whom welfare is *for*, but this person must therefore be able to value or disvalue things which affect *their* welfare; that is, they must be a 'valuer'. As value cannot exist without valuers (in this case, welfare cannot exist without being *for someone*), then 'valuers have the unique status as the *condition* of there being any value'.¹⁵ In the cases under discussion (prospective parents seeking fertility treatment), the only valuers for whom the decision can be of value are the parents. The potential child therefore cannot have a 'welfare concern' that can be affected when it is the decision which will determine whether it exists or not.

This seems clear and highlights the fact that deciding against the existence of potential person cannot count as beneficial *to them*. Obviously it could be beneficial to the parents as existing valuers, because they would not have the burden of raising a (disabled or non-disabled) child, and it could be beneficial for other actual members of society, for example, due to the consumption or distribution of health care resources. But it is painfully obvious that it cannot be beneficial to the potential person when they will never exist. For example, genetically modifying an embryo so that it no longer has a genetic disability would not benefit it, because the decision to genetically modify would constitute the potential person; that is, it would be a decision prior to and determinative of their existence. There is no basis for the claim that offspring are harmed by genetic modification when it is *that genetic modification* which causes them to exist.

Furthermore, 'equally and for the same reasons, we cannot hold the child to be an object of maleficence'.¹⁶ Just as a valuer needs to exist before they can benefit from things so they have to exist before they can be harmed by things. For example, a potential disabled child cannot be harmed by being brought into existence. Thus 'potential people have equal standing regarding our beneficence and our nonmaleficence',¹⁷ which is to

15. Heyd, 'People', p. 117.

16. Heyd, 'People', p. 109.

17. Heyd, 'People', p. 115.

say they have *no* standing. Therefore there is no distinction between obligations to benefit or to not harm a potential person; neither type of obligation is engaged. However, this also means that creating new valuers is itself of no value *independent* of those valuers who already exist, there thus can never be any presumption to implant an (or any) embryo on the grounds of benefitting the potential person created by this process. This also means that there can never be a presumption against implanting an (or any) embryo justified by avoiding harm to the potential person. It is only on the basis of the interests of actual persons, the parents, that decisions regarding offspring can be made.

Once the potential person does exist then they can be affected by the circumstances in which they exist. The necessary precondition for value (harm and benefit) to attach to someone has been met (by their existence) and from there onwards they possess interests because they begin to form a 'biographical life'. A biographical life consists of the actions, desires, projects and decisions that make up a life¹⁸; that is, the components that give a life consistency. It should thus be self-evident that an existing child has interests that can be accounted for and protected, and these post-existence effects can form the basis of welfare decisions. At this point, the Children Act 1989 section 1 would be applicable and would protect the actually existing child, further eroding the purpose and need for the welfare test in the HFE Act. The case of disability is a good example of the argument outlined above.

If an embryo has the genetic characteristics of being disabled, the HFE Act considers the disability to count as a reason for rejecting that embryo and prohibits selecting that embryo when a non-disabled embryo is available.¹⁹ This is so even though the offspring would not actually come into existence because the embryo has been rejected and thus no subject would ever exist to benefit or suffer harm from the decision taken. However, this also applies to a decision to allow the disabled embryo to develop to term, because the potential person must exist *before* the 'person-affecting' nature of value can be present. In such cases (rejection, acceptance or modification of the disabled embryo), the potential person cannot be affected by the decision taken, but the parents can. Thus, it is their interests which are relevant when deciding which embryo to implant and it is up to them to determine whether having offspring with a disability is against or in their interests. Once a child exists (whether disabled or not) then the precondition for value to be *for them* has been met and they will possess interests that can be benefited or harmed. The welfare of the child test under the Children Act will now apply and judgements regarding the child's welfare can now be made, thus forming the basis for decisions and actions about the child's upbringing and care.

As noted throughout this article, the basis of removing the potential person (the child who may result) from consideration is the necessity of their existence before they can be harmed or benefited. This applies equally to negative constraints (such as not selecting certain embryos) and positive requirements (such as genetically modifying an embryo). Consequently, it is impossible to maintain the HFE concept of the welfare of the

18. Statement by John Harris (personal communication, 12 April 2013).

19. Human Fertilisation and Embryology Act 1990 as amended, *Human Fertilisation and Embryology Authority Code of Practice*, 8th ed., 2009.

offspring who results from the treatment because potential people have no welfare to speak of. The welfare of the (potential) child test should therefore be abolished and the consideration of the potential child removed as a factor for consideration when deciding on distributing reproductive technology and services. Any restrictions or constraints imposed upon parents would then have to be based upon the impact on actual persons which would include the parents seeking reproductive aid and the other actual members of their sociopolitical unit (more on this later). The general form of the test found in the Children Act, with its focus on existing offspring, would, of course, retain its applicability to offspring once they exist.

We can thus see how this change would eliminate the restriction on parents when it comes to the selection of different embryos for implantation. Next, we will consider an alternative to the person-affecting approach, that is, impersonalism. If there is an alternative to the claim that welfare must be person affecting, then this may undermine the claim that welfare must be *for someone*. This alternative, however, can only be based on a non-person-affecting account, namely, impersonalism.

Impersonalism: The alternative to the person-affecting approach

As has been made clear throughout this article, the welfare of the (potential) child is logically unsustainable if welfare is attached to a subject. Thus, if welfare can be framed in such a way that it does not attach to a subject, this would offer another method for regulation to incorporate the welfare test. To be clear, impersonalism here is not 'a common standard'²⁰ that is good *for everyone*; as this would simply be claiming that there is some *x* which is good *for all subjects* who exist. It would thus still be tied to the actual existence of people but would simply be a claim that *x* is good for all of them. This obviously would not be an alternative to, or an answer to, the argument set out above because it would still require that a person exist for them to benefit from the common standard. Here, impersonalism is used to mean the existence of value as a feature of the world; that is, value is independent of persons and attaches 'to the *world*'.²¹ This stance would thus identify what is good and claim that the more of this good, the better because it is as a feature of the world that it has value, not as it relates to persons (if they exist).

So what might an impersonalist approach look like? Julian Savulescu argues for an approach in which the offspring 'who is expected to have the best life, or at least as a good as life as the others'²² should be chosen. We would therefore not select a disabled foetus because it would not have the best expected life. However, as Rebecca Bennett puts it, 'This individual is born in this impaired state or not at all',²³ and thus the choice is between the existences of different potential people and on a person-affecting account

20. Heyd, 'People', p. 82.

21. Heyd, 'People', p. 80.

22. J. Savulescu, 'Procreative Beneficence: Why We Should Select the Best Children', *Bioethics* 413 (2001), p. 413.

23. R. Bennett, 'The Fallacy of the Principle of Procreative Beneficence', *Bioethics* 263 (2009), p. 267.

this raises the same problems outlined above. Consequently, we could not say it is better *for those who will not exist*, as this would be to take a person-affecting perspective. Of course, on a person-affecting view, we could say that it is better for those of us who actually exist that this particular potential person does not exist. For Savulescu's comparison to work, it would have to be based upon impersonal criteria, particularly as the existence of each potential person is mutually exclusive of the others. In order to avoid relying on the effect on *any* people, what we would have to say is that the world is a better place with those particular potential persons in it, rather than these other potential persons, because those we allow to exist will contribute more good (in this case the best lives) to the state of the world. In other words, the world is better off than with alternative persons in it because they add to the total good in the world. Alternatively, it could be that impersonalism is based upon being the right kind of person (i.e. being human consists of suffering and limitation).²⁴

The essential problem with impersonalism is that it 'sacrifices the utility of individuals to the promotion of the impersonal value of the overall good'.²⁵ In other words, whatever the good is, people are merely a means to produce the good, and thus individuals are always expendable. This means that people are not relevant beyond the production of the good. For example, if health was the primary good (as suggested in Savulescu's scenario), then all parents could be forced to undergo whatever procedure maximized health, as long as the cost in the health of the parents was outweighed by the gain in the offspring. If impersonalism is based upon creating the right kind of people then this balancing act would be unnecessary; it would only be when the wrong kind of offspring were to be produced that parents would have to take measures to prevent it *whatever those measures are*. Now it is obvious that if impersonalism is the correct approach, then the subjugation of individuals to the production of the good or to the creation of the right kind of person is not only permissible but an obligation. The restrictions imposed upon individuals could thus be justified.

But impersonalism is not a sound or coherent theory because the identification of the good or the right kind of people unavoidably *presupposes* the existence of valuers. The good that forms the basis of impersonalism (whatever it is) is 'by necessity of actual'²⁶ persons. Without actual persons, that is valuers, it is impossible to differentiate between different conceptions of the good because a world without persons is valueless. Otherwise anything could be good, even the murder of humans could be justified on the basis that it is *more natural* for a predator species to behave in such a manner and thus promoting the good of the natural world. Thus, the existence of persons who can identify the good is a prerequisite for impersonalism to work, but this is the very element that impersonalism claims to do without. More to the point, impersonalism itself gives all individuals a reason to reject it because it could impose upon them a good which *they* do not find to be good. Impersonalism can therefore only work at one level of discourse, namely the

24. L. Kass, 'Ageless Bodies, Happy Souls: Biotechnology and the Pursuit of Perfection', *The New Atlantis* 1 (2003), p. 9.

25. Heyd, 'People', p. 57.

26. Heyd, 'People', p. 63.

identification of objective goods *for a particular reference group*. It cannot be used to justify the existence of members of that reference group because the identification of value *without people* is impossible. Furthermore, the identification of a good to which persons are subservient is not possible, because good has to relate to the person or persons affected.

Thus we cannot resort to impersonalism to solve the conundrum of the potential child possessing welfare interests, because impersonalism reduces persons to mere producers of a good and thus justifies any restriction or interference which promotes this good. It is also conceptually incoherent as the good cannot be identified without presupposing the existence of the reference group (valuers). Once the reference group exists, then the conception of the good becomes person affecting, as only by tying good to valuers does it make sense to differentiate between potential goods. The arguments pertaining to the illogical and impossible conception of welfare *for a potential person* remain valid and cannot be replaced with impersonalism. We therefore have to accept that the welfare of the potential child test is unsustainable and that the law is wrong to have such a test because it is factually impossible for the test to have *any effect* on the person it supposedly protects. The only option left is to remove the test completely. This does not however mean that parents have or should have unfettered decision-making powers over reproduction, but it does mean that such restrictions must be based upon the interests of actual persons; in this case, the parents and the other members of their sociopolitical unit. It is to this consideration that we now turn.

The decisions of prospective parents

Having dispensed with the welfare of the potential child test, we need to decide if and on what basis the reproductive decisions of prospective parents could be constrained. As the preceding section pointed out, without the welfare test and the requirement that the decisions take a person-affecting form, reproductive 'genesis choices . . . should be guided exclusively by reference to the interests, welfare, ideals, rights and duties of those making the choice'.²⁷ Those who are making the choices are the 'generators' of the (pro)-creative effort and, as it is their interests that are of concern, Heyd calls this theory of their predominance 'generocentrism'.²⁸ Generocentric theory will form the foundation of the new reformed approach to reproduction, specifically in the case of parents utilizing reproductive technology and services, where the decisions of the parents are contingent upon the acceptance of those providing the service for the decisions to take affect. Artificial reproduction therefore provides a case in point where the interests of other actual persons and the interests of parents intersect.

From the arguments set out throughout this article culminating in generocentric theory, there can be no restrictions imposed upon parents from the perspective of the potential child. Parents should therefore have greater control over the reproductive process unconstrained by the supposed impact on the potential child. What is of interest here

27. Heyd, 'People', p. 96.

28. Heyd, 'People', p. 96.

is what possible constraints there might be for the parents in the absence of the welfare of a potential child. As pointed out earlier, any such constraints would have to be based upon either the interests of other people or upon a requirement for the parents to meet a standard of conduct. Given the emphasis on the person-affecting nature of actions, an approach grounded upon the interests of other people is the only justification for restricting the actions of parents, but this could of course take the form of a standard of conduct. However, we will avoid becoming encumbered by discussions between the relative merits of the different forms the restrictions may take. We shall instead focus on the effects that would justify imposing the constraints.

Some examples are the increase of health care costs; ensuring genetic diversity for the gene pool; preventing hyperaggression; improving immune systems, muscle strength, liver function, and brain functioning and so on. The justification for restraints could also relate to the kind of society that we are trying to build, whether it is strongly democratic and pluralistic with a strong sense of individualism, whether strict conformity to a particular ideal is required or whether reproduction is considered a collective effort or interest. Given that the jurisdiction under discussion is the United Kingdom, this article takes it as given that the constraints have to be compatible with a democratic pluralist human rights respecting society. In this case, the basis of restrictions has to be proven by those seeking to restrict the actions of the parents. This would be even stricter if a formal right to reproduction or reproductive autonomy were to be created or recognized, because the restriction would also have to be necessary and proportionate. Unfortunately, the decision as to whether there should be such a right is beyond the scope of this article.

Thus, it must be noted that this person-affecting approach would entail a strong form of reproductive autonomy for parents generally in the area of reproduction as the impact upon others would have to be proven. For example, if someone claims that it would be harmful to have a disabled child then they must show to what interests it is harmful; such a claim could not be based upon simply disapproving of the choice. It is therefore plausible that only a significant interest could restrict parental decisions. Such a significant interest might be the interest of actual persons to have access to health care. Thus, parents who choose to have disabled offspring might be constrained on the basis that the additional resource burden on the health care system and the subsequent effect on extant people's interests make the choice to have disabled offspring an unreasonable expectation for the use of public resources. It could not, however, be based upon the harm done to the parents themselves, because it is up to them whether the burden of raising disabled offspring is outweighed by the benefits of having a child, but it could easily be based upon the effect on others. Such a justification would encompass concerns of fair distribution of resources, safeguarding and balancing conflicting interests (and/or rights) among a population, population growth policy and societal aims, changes or objectives. One scenario may be that the interests of actual persons make it preferable to ensure a universally accessible standard of reproductive technology rather than a private system that only some could afford.

Other modifications however may be problematic; skin colour, for example, may be something that remains outside of parental remit, not because it may create a burden or thwart the interests of others but because allowing such a choice in a society which unjustifiably discriminates against certain skin colours may cause reproduction (and

genetic modification) to be used to reinforce racism and other prejudices. Such a consequence could be opposed because diversity is valued by society as leading to a richer and deeper culture, because it is a superficial difference or because the promotion of a united human governmental system or human rights culture are considered important for improving the lives of actual persons. In this case, the justification could be based upon the harm done to actual persons in the sense that they would live in a society that is worse *for them* because it violates the basic principles or objective of society or because it worsens the situation of actual persons. Deciding on what may and may not be permitted is up to the sociopolitical unit where artificial reproduction takes place; it is beyond the scope of this article to consider the political methods which should be used. However, in no case could it be based upon the potential person themselves, as it is impossible for them to have interests and because they are irrelevant themselves except for their value to extant persons.

Finally, imposing constraints on parents may be justified because we (i.e. the rest of society) *do not want them* to achieve or satisfy their interests. For example, parents who want to inflict suffering and harm on a person may be prohibited from utilizing reproductive technology and services because we do not want to help people satisfy their desires to act in cruel and abusive ways by creating a child. In such cases, we are not saying that the child's interests would be harmed by being created for such a purpose, and once they exist, their interests would justify removing the child from the parents, although this may be a difficult and protracted process. Rather, we are saying that the interest the parents seek to fulfil is *prohibited* because society seeks to minimize the prevalence of cruel behaviour among actual persons; thus they cannot be permitted to achieve their goal. This would be similar to not allowing people to vandalize public or abandoned property, to torture willing 'victims' or to take class A drugs even if no one is directly harmed by such conduct. This is because some conduct is deemed unacceptable by society and those who would use reproductive technology and services to fulfil these prohibited objectives could be prevented from doing so. Obviously these prohibitions purport to represent more than just likes and dislikes of society; how we might determine which rules to adopt would have to be underpinned by a full a normative theory.²⁹

Regardless, within the United Kingdom and other democratic pluralist human rights respecting nations, the interests and decisions of the parents would have priority. This approach would therefore create a legal presumption or right in favour of permitting prospective parents to reproduce the offspring they wish, and those who wish to restrict the actions of parents would have to make their case and provide evidence of their claims. Thus the person-affecting approach does not preclude restricting the actions of prospective parents on the basis of the interests of other persons or on the basis that prospective parents should act in certain ways. This would be based on competing interests between societies, existing members and the parents. This provides the justification for judicial oversight of reproductive technology whilst protecting parents from being denied access to reproductive services on the basis of the welfare of the potential child.

29. For those who wish to consider this point further, contractarian theory provides a good example of developing rules based upon a normative system of rational agreement.

Conclusion

The fundamental argument of this article is that potential persons have no standing (moral, ethical or legal) in relation to those who actually exist. Consider the following, imagine that every person on the planet joins the Voluntary Human Extinction Movement³⁰ and through a program of permanent and irreversible sterilization the human species ensures that it cannot reproduce by any means. In such a case, all potential persons would cease to be potential at all, because *none* of them will ever exist. We could not say that we owe a duty to these non-existent beings of a no longer possible future; and if we do not owe an obligation to them to ensure they exist, then we cannot owe them any duty to have a *particular* life. This also has implications for environmental strategies, as it would mean that we cannot owe them a duty to provide any particular environment or to provide as certain level of resources. A potential person cannot benefit from existing or not and neither can a potential person be harmed nor are they owed any obligations.

Consequently, when it comes to reproduction, the interests of prospective parents are unconstrained by the effect of their actions on their potential offspring because actions which cause a person to exist cannot affect them. Parents are thus free to choose to have disabled, genetically modified or 'normal' offspring as they wish. The only source of restrictions on parental conduct comes from the effect of their actions on other actual persons. This approach is explicitly person affecting which has as a precondition for it to apply that a person *actually* exists. It is this that underpins the argument that potential persons have no standing compared to the interests of actual persons, be they parents or members of society.

The importance of this change in outlook is prompted by the inclusion of the welfare of the (potential) child test in decisions about access to reproductive services by infertile couples. This is objectionable because those already experiencing difficulties in having children are further penalized by having to prove their capacity to be parents; and all this is because of a potential person who cannot be affected by their actions. The point at which the potential offspring becomes actual (when the child can be affected) is covered by the Children Act 1989. Thus, parents have their choices, decisions and actions restricted by the interests of a being that does not exist. When we consider the impossibility of harming potential persons, we can see that this is an unjustified restriction imposed upon parents which consequently should be removed. Only by taking an impersonalist non-person-affecting approach could we avoid this difficulty.

Impersonalism has however been shown to be incoherent and flawed because it is abhorrent to individuals by making them into producers of value (reducing their own interests into irrelevance) and because it presupposes the existence of valuers in order to determine and identify the value(s) that are to be produced. Note once again that this is not to say that impersonal (objective) considerations are irrelevant when considering benefits across an identified reference group but that the very foundation of benefit, harm and value cannot be disconnected from the existence of evaluators. Thus impersonalism needs persons to exist before it can have any meaningful content, but it then discounts the interests those persons have once they do exist. The alternative to the person-affecting

30. Available at: <http://www.vhemt.org/> (accessed 29 August 2014).

approach is therefore untenable, and this brings us full circle back to the determination of the welfare of the potential child test as unacceptable and leaves us with only the interests of actual persons to consider.

Thus parental decisions would be unrestricted by the concept of harming their future offspring as those actions are the ones that bring the offspring into existence. We should note however that it has recently been shown that there is a general and widespread uncritical acceptance of the requirement and that the regulatory regime operates in a 'light touch' manner.³¹ This suggests that because there is such widespread belief, enforcement can be light touch. Without this rule, as proposed by this work, then people may still act on the basis of concern for the welfare of future persons. This suggests that the existence of the rule may be unnecessary and could be interpreted as supporting my position. On the other hand, it could be argued that if such belief is widespread then it should be reflected in the legislation. Either way it calls for further empirical analysis of service users and providers experiences in artificial reproduction.

Parents would thus be able to utilize reproductive technology (including genetic modification) with only legitimate public interest limiting their actions. Of course, medical providers will have obligations to the parents, and parental actions will be restricted by the technological capabilities of the time. Moreover, depending on how services are provided, for example, by a public or private system, parental decisions will be limited by different constraints. What is made clear is that parental interests have primacy and can only be overridden by the interests of other existing citizens and not by the interests of the potential offspring. This could lead to a legal presumption that parental decisions to have disabled offspring or genetically modify their offspring are in their interests and that if others want to prevent such modification going ahead, then the burden is on them to prove that there is a sufficient public interest or harm being caused to actual persons.

As we have seen, this may take a number of forms but can be based on the harms caused directly to other persons, the fairness and distribution of resources or on the ideals a society holds. For example, we want to prevent those who sadistically want to torture people from achieving their aims, or we want to ensure that every member of society has access to services on an equal basis and so some limitations to the number of times reproductive services are accessed is imposed. Another example may be prohibiting the selection of skin colour because we do not want to exist in a more prejudiced society than we do now. All of these examples, whether to do with the harm caused to a specific individual, the fairness of health care access across the entire population, population control policy or because of the particular ideals a society holds, are based upon the impact and effect of actions on existing actual persons. Restrictions and constraints imposed on this basis are therefore permissible and justifiable. However, depending on what form the parents' interests take, for example human rights, will affect the strength needed for justification and the degree to which interference can extend.

31. See E. Lee, J. Macvarish and S. Sheldon, 'Assessing Child Welfare Under the Human Fertilisation and Embryology Act 2008: A Case Study in Medicalisation?', *Sociology of Health and Illness* 36(4) (2014), pp. 500–515.

The reform recommended here, that is, the removal of the welfare of the (potential) child test in relation to artificial reproductive services, would clearly make the interests of the prospective parents paramount. As the parents are the ones whose interests would be fulfilled by reproduction and because it is such a major component of their lives, it is fair that decision is accepted *unless* it can be shown that their decision is harmful to an actual person or because it violates the standards of conduct held in the sociopolitical unit. A decision as to whether parental discretion should be embodied in a human or legal right is a discussion for another time. It is clear, however, that the injustice and unfairness imposed by the welfare of the (potential) child test is untenable and should be excised as part of the process of artificial reproduction. If the actions of prospective parents do not harm anyone or violate behavioural codes then they should be free to reproduce as they wish.

Applying the actual/ potential person distinction to reproductive torts

Medical Law International
1–19

© The Author(s) 2014

Reprints and permission:

sagepub.co.uk/journalsPermissions.nav

DOI: 10.1177/0968533214537288

mli.sagepub.com



Samuel E.P. Walker

University of Manchester, UK

Abstract

As technology has advanced, the level of control that can be exercised over the reproductive process has increased. These advances have resulted in a number of claims in tort law relating to pregnancy and birth. The three reproductive torts considered here are ‘wrongful conception’, ‘wrongful birth’ and ‘wrongful life’. This article will consider the theoretical underpinnings upon which these torts rest and will suggest that the potential/actual person distinction is crucial to these reproductive torts because potential persons should not be able to make claims in tort based on alternative conditions that could never have been. This is because actions (or omissions) prior to birth determine the preconditions for existence. Thus, only actual persons (i.e. those who exist at the time of the action or omission) should be able to bring claims in tort. The analysis will conclude by arguing that no child should be permitted to bring a claim under any form of reproductive tort. The term reproductive torts originates with Nicolette Priaulx’s work and encompasses all three terms: wrongful life, wrongful birth and wrongful conception. While these are distinct terms and all fall within negligence for the purposes of this analysis ‘reproductivetorts’ is a useful term to identify these particular claims.

Keywords

Potential persons, tort, reproduction, wrongful life, wrongful birth, wrongful conception

Corresponding author:

Samuel E.P. Walker, School of Law, University of Manchester, Manchester M13 9PL, UK.

Email: sep.walker@gmail.com

Introduction

The main purpose of this work is to consider the implications of acknowledging the potential/actual person distinction in relation to legal claims relating to reproductive torts. This entails consideration of what constitutes a 'person' and what impact decisions in this area will have upon the existence and identity of potential persons. The discussion in this article will be anchored around a discussion of the reproductive torts of wrongful birth/conception and wrongful life claims. For the sake of simplicity, wrongful birth and wrongful conception will be discussed together because they both relate to parental claims rather than children claiming. These torts have been selected because the very basis of these claims relates to the existence and identity of a person *who would not have existed but for the action giving rise to a tort claim*. Collectively these torts deal with actions that cause an individual to exist. Furthermore, the reproductive torts of wrongful conception, birth and life are pertinent to discussing liability for reproductive actions because they intersect with claims between parents, offspring and third-party medical providers. Reproductive torts thus provide a suitable context within which to discuss the potential/actual distinction, psychological personhood and the implications for legal reform.

The format of this article will be as follows. Firstly, it will provide an outline of why the potential/actual person distinction is important and what that implies for how we think of decisions that create new individuals. This will be followed by a section assessing the construct of psychological personhood, which is important because it highlights that changes occurring *before* a potential person becomes an actual person create new individuals rather than changing the conditions for a pre-existing one. This argument is based upon the fact that different psychological characteristics create different people and that this is affected by both mental and physical changes to an organism. Thus, prior to actual existence, no person exists and changes to the conditions in which offspring will be produced are creating new persons. Together these two related concepts of potential/actual persons and psychological personhood will serve as a platform for reassessing reproductive torts and in suggesting the direction reform of the law should take. Essentially the claims will be as follows: potential persons have no claims on the actions of actual persons and only the interests of actual persons can sustain claims in tort.

As we consider the reproductive torts of wrongful birth/conception and wrongful life we will see how the courts have encountered problems, which the theoretical section on potential/actual persons and psychological personhood expose. In further support of the integral nature of the problems with reproductive torts, other non-UK jurisdictions will be brought into the discussion. For the sake of clarity, the differences between the reproductive torts under consideration will be explored in the brief description of each 'tort' that follows.

'Wrongful conception'

Wrongful conception cases concern the birth of a healthy child where the very conception of a child was something that the parents had actively sought to prevent (e.g. by undergoing a vasectomy which was carried out incorrectly, so that the parents now have

a child which they never intended to have.¹) ‘Wrongful birth’ cases occur when a child has been conceived and the parents are given incorrect advice or information regarding the condition of the foetus, where such advice if given non-negligently might have led to a termination.² These cases also include situations where some action by the medical professional causes a disability to occur or damages the foetus.³ Normally these reproductive torts are relatively straightforward claims by existing persons (the parents) that they have been harmed because the medical procedure was carried out incorrectly and resulted in the birth of a child, with or without a disability.

However, in the United Kingdom, there is an aberrant situation under section 1 of the Congenital Disabilities (Civil Liabilities) Act 1976 (‘CDCLA’), which allows a child to claim under the tort of wrongful birth when the actions of a medical provider lead to the child having disabilities that it otherwise would not have. It is important to note that this does not allow claims by healthy children; it only allows claims under wrongful birth in cases of disability caused by medical providers. In this respect, wrongful birth claims under the CDCLA are somewhat similar to the wrongful life cases.

‘Wrongful life’

The tort of wrongful life is a claim by a child that the child should not have been caused to exist at all. In essence, offspring claim that life is so intolerable and full of inescapable suffering that *they are better off not existing* rather than being born in such a state. The courts are thus asked to conclude that being born was harmful to the child and that non-existence was a better condition, of which they have been deprived. The term wrongful life is not without controversy as Kirby in the Australian case of *Harriton v. Stephens* demonstrates. Kirby stated that in his view ‘its use, even as a shorthand phrase, should be avoided’.⁴ Kirby takes this view for a number of reasons, but the most important of them are as follows: the phrase wrongful life ‘denigrate[s] the value of human existence’ and that by ‘lumping all such cases under the one description there is a danger that important factual distinctions will be overlooked or obscured’.⁵ Kirby prefers the term ‘wrongful suffering’ as it encapsulates, what Kirby sees as the basis of the claim, that the ‘negligence . . . has directly resulted in the present suffering’.⁶ In line with judicial terminology, the phrase wrongful life will be retained in this article, although as we will see, in addition to the just stated objection, this term has different interpretations in different jurisdictions.

However, this raises the philosophical problem of comparing existence with non-existence, which is conceptually impossible because non-existence is the absence of everything. There is thus *no one* who is better off than the child that does exist. This is also a legal problem for the courts as they cannot (and are not prepared to) calculate

-
1. *MacFarlane v Tayside Health Board* [2000] 2 AC 59.
 2. *Farraj v King’s Healthcare NHS Trust* [2010] 1 W.L.R. 2139.
 3. *Whitehouse v Jordan* [1981] W.L.R. 246.
 4. *Harriton v Stephens* [2006] Unreported Cases High Court of Australia at [8].
 5. *Harriton* [2006] at [11].
 6. *Harriton* [2006] at [10].

damages on this basis and are unwilling to countenance the idea that non-existence could be preferable to life.⁷ Currently, such claims are not permitted under UK law⁸ and, as we shall see, other jurisdictions have struggled with the wrongful life tort. In the New Jersey case of *Gleitman v. Cosgrove* (1967),⁹ the court found that a wrongful life claim could not be actionable, while in the state of California the case of *Curlender v. Biosciences Laboratories* (1980)¹⁰ held that a wrongful life claim was sustainable. It is thus by no means a settled question as to whether wrongful life claims can be judicially recognized.

Moreover, wrongful birth and wrongful life actions can be distinguished as claims brought by parents and claims brought by the offspring themselves, respectively,¹¹ however, this division on the basis of the identity of the claimant does not always apply, as we have seen from the CDCLA. Throughout this article, these reproductive torts are separated by type (i.e. wrongful birth/conception and wrongful life) rather than on the basis of who brings the claim. It should, however, be borne in mind that who can claim, normally, follows the pattern of 'wrongful conception/birth', the claimants being the parents, while wrongful life claimants are the children (or someone acting on behalf of the child); but this is not always the case and sometimes the claimants will be different as under the CDCLA.

Reproductive torts relate to harms and damages that the law recognizes to offspring where the action has occurred before their birth and covers both parents and the child claiming against a medical provider. The law can therefore recognize that harms can occur before a legal person comes into existence and that once the child exists his/her claim crystallizes, allowing a claim in tort. However, the conclusion of the article will be that children can never claim for the condition they are born in, regardless of what state that is or how it was caused. This would require the removal of section 1 of the Congenital Disabilities Act, which allows children to claim under wrongful birth cases, and retain the non-recognition of wrongful life torts. This conclusion will be based upon the fact that decisions and conditions which constitute the new person cannot be grounds for claims that the person either should have existed in a different condition or should not exist at all, because in both cases the person (the person claiming) would not exist and thus the person would not be affected. This line of reasoning is 'person-affecting' in its approach, which dovetails with tort law as tort requires that *a person* is in a worse condition than they otherwise would be; that is, a person is affected by the wrongful conduct.¹²

This conclusion would apply whether the conduct comes from medical providers, prospective parents or genetic engineers (whether through genetic selection or modification), thus only allowing for the interests of actual persons to determine the decisions made. Therefore, the legal claims of reproductive torts will not be able to be based upon

7. *McKay v Essex Area Health Authority* [1982] Q.B. 1166.

8. *McKay* [1982].

9. *Gleitman v Cosgrove* [1967] 49 N.J. 22.

10. *Curlender v Biosciences Laboratories* [1980] 165 Cal. Rptr. 477.

11. *Harriton* at [1].

12. C. Elliott and F. Quinn, *Tort Law* (Longman, 2011), p. 375.

harm to the not-yet-existing offspring but would have to be based upon the effects of their actions on actual persons.

The potential/actual distinction¹³ and psychological personhood

The importance of the potential/actual person distinction cannot be overstated because it speaks to the difference between those who can suffer harm or gain some benefit and those whose very existence is constituted by the decisions being undertaken. It is important because the basic aim of tort law, including reproductive torts, is to put someone in the situation they would have been in if the tortious action had not occurred.¹⁴ The problem with reproductive torts is that the situation before the tort occurred is non-existence, and thus a comparison between existence and non-existence is required to calculate damages. As Ackner states, 'how can the court begin to evaluate non-existence. . . . No comparison is possible and therefore no damage can be established which a court could recognise' (*MacKay v. Essex Area Healthboard*, 1982 at p. 1189). Thus, we do not have to consider whether we take a person-affecting or impersonal approach to harm, because tort presupposes that persons *are* affected. We therefore must be able to identify the person who is affected and we *must* know how they were affected and what condition they would have been in prior to the action.

We can thus see why the potential/actual person distinction is important. Decisions which create persons cannot be subject to considerations of harm and benefit in relation to the person yet to exist, because the very conditions for affecting persons require that they actually exist and potential persons do not fulfil this requirement. Moreover if, as in the case of reproductive torts, the action which is complained of did not occur, then the person would never have existed and thus there would be no person in a better or worse condition. As David Heyd puts it, the 'confusion created by wrongful life cases arises precisely from the fact that the wrongful act is the direct cause of the plaintiff's existence'.¹⁵ Although Heyd is speaking only in relation to the tort of wrongful life, this article will argue that the person-affecting requirement of tort is never satisfied in reproductive cases when the child brings the claim. This is because potential persons do not exist, and as such changes to the embryo's genetic make up or physical body do not change *anyone*; rather they alter the genesis conditions, which create a person. For example, changing whether an embryo has a disability will not change the (potential) person who will exist because no one exists, it will only change the conditions of the creation process.

Additionally, the existential status of potential persons means that they also fail to fulfil the need for a single relationship. As Ernest Weinrib states tort law, as a component of corrective justice, 'treats the doer and the sufferer of harm as the active and passive

13. The potential/actual distinction is a concept that arises in across the literature. See D. Heyd, *Genethics* (Berkeley: University of California Press, 1992); A. Giubilini and F. Minerva 'After-Birth Abortion: Why the Baby Should Live?', *Journal of Medical Ethics* 39(5) (2012), p. 261.

14. Elliott and Quinn, 'Law', p. 375.

15. D. Heyd, 'Genethics', p. 29.

participants in a single relationship'.¹⁶ Thus, the legal relationship required in tort claims is absent because of the necessary requirement of existence for there to be a doer and sufferer, a harm and a harmed.

What are psychological persons?

This article adopts a psychological personhood approach to the nature of persons and personal identity, which will now be outlined. The concept of psychological personhood is a claim that can be summarized as, persons exist only as an aggregate or a construct of the mental capacities of organisms, they do not exist outside of this realm. Thus, a person is not a person because the person is biologically human, or because the person possesses a name, but because the person has certain psychological features of memory, self-awareness, a sense of self over time, desires and interests between the past, the ongoing now and future instants.¹⁷ These features then give rise to psychological *connectedness* (connections between one instant and the next) and together form psychological *continuity* between past, present and future.¹⁸ Lynne Baker identifies the crucial attribute of a person as the 'first-person perspective' (FPP) in which a being can think of itself as the source of thoughts and desires.¹⁹ This is a good means of identifying a person and provides a continuous sense of identity within the entity that is a person. The content of personal identity, that is, of a person being a particular someone and having specific goals and desires is provided by the other psychological features of experience and memory. Thus, the FPP provides a sense of unity over life and experience, but FPP relies on connections and continuity between instants and experiences to provide personal identity. Personal identity is therefore the narrative of a specific person's existence, while being a member of the group of beings called person is dependent on the capacity for an FPP. If these connections are absent, then psychological personhood would suggest that someone is not the same person over time even if the person remains the same physical organism. This allows for a self-referencing narrative, which builds a biographical life, gives meaning and purpose of a person's life and makes a life plan constructed through projects²⁰ possible; this self-recursive construction is what distinguishes a person from a non-person.

This starkly contrasts with how we normally think of ourselves. As a consequence, the importance of psychological attributes is greater than under alternative theories of persons (such as the view of persons being an indestructible soul or being a human organism). Furthermore, it means that persons are fundamentally the font of value and for things to be 'good' or 'bad', a person must *actually* exist. An actual person who exists

16. E. Weinrib, 'The Jurisprudence of Legal Formalism', *Harvard Journal of Law and Public Policy* 16 (1993), pp. 583, 588.

17. J. Locke, *An Essay on Human Understanding* (Menston: Scolar Press, 1690); D. Parfit, *Reasons and Persons* (Oxford: Clarendon Press, 1984), pp. 204–209.

18. D. Parfit, 'Persons', pp. 204–209.

19. L. Baker, *Persons and Bodies* (Cambridge: Cambridge University Press, 2000), p. 64.

20. L. Lomasky, *Persons, Rights and the Moral Community* (Oxford: Oxford University Press, 1990), p. 32.

can be affected by things that happen around and to the person and thus can make a value judgement regarding those actions. Potential persons cannot be affected in a similar manner because the necessary condition for them to evaluate the effect of things is that they exist, but it is their very existence which is being decided.

Thus, Heyd can state that 'they [potential persons] do not meet certain preconditions of existence and identity',²¹ and consequently as we modify and change an embryo (i.e. as we change the conditions for the potential person's existence) we will be creating a new person, not altering the circumstances for the same person. If this were not the case, for example, if persons were 'souls', then reproductive torts would be sustainable, as the existence of the person would be prior to the existence of his/her body. However, in this case, every negative attribute would be included in a tort claim because these would all be *affecting someone* and would be making things worse for them. In the case of psychological persons, the person develops after his/her body, thus actions which occur prior to the person's existence are excluded from *affecting* them because these actions constitute the person who will exist.

This theoretical platform means that actions giving rise to reproductive torts will lead to a new person coming into existence regardless of what these actions are. Thus, medical providers causing a disability to occur in an embryo will simply be changing the constituting conditions of a new potential person. Similarly wrongful conception results in a new potential person who would not otherwise exist. Equally, failing to diagnose a disability that would lead to the parents terminating the pregnancy would also be creating a new person. In both scenarios, no one is being made better or worse off, no one is being replaced with a different person; rather each scenario creates a new person, generating a new personal identity from the specific constituting conditions. As wrongful life claims would have seen an actual person not come to exist, potential persons have no claim upon actual persons and wrongful life claims must be rejected. But we will now consider opposing views to this theory.

Potential persons and the 'right' to be born (healthy)

Having set out the argument for reforming reproductive torts by eliminating claims made by the child himself/herself, it is only fair that some counterclaims arguing that potential persons have interests are considered. The obstacle that must be overcome in order for potential persons to have any interests or rights over actual persons is the potential/actual distinction, because it is this divide that precludes claims based upon actions that affect the constituting conditions of a person's existence.

So how then can interests and rights translate across the potential/actual divide? Loren Lomasky suggests that harms that affect a future person, but which are caused before the person exists, can give rise to a retroactive complaint. This is because actual persons can recognize 'goods-for-the-child [or potential person] that others have reason to acknowledge and respect',²² and recognize that the future person is 'identifiable as a distinct

21. Heyd, 'Genethics', p. 36.

22. Lomasky, 'Community', p. 161.

individual upon whom one can act for better or ill'.²³ Walter Glannon suggests that we can 'prevent actual (future) people from experiencing pain and suffering and thereby avoid defeating their interests in having healthy lives'.²⁴ Moreover, Jonathan Glover states that 'the parental desire for a healthy child fits with the interests of the child born as a result of the choice'.²⁵ These statements suggest that we can both identify the person who will exist and extrapolate from that things which are/will be good for them.

Lomasky claims that recognizing 'good-for-someone' and being able to identify them provide a way to identify harms (over which we have control) for future persons, thus we should act to prevent these harms occurring. In other words, because offspring will become persons *who are now identifiable* and we can predict the impact of our actions on these future persons we are subject to their interests. Similarly Glover and Glannon claim that a potential person has an interest in being born in a healthy condition. It thus seems that, to Glover at least, once a conception occurs we can identify some interests of the future person. Glover names, for example, an interest in being 'able-bodied and healthy'.²⁶ Therefore, 'the interests of the child should set limits to what potential parents do'.²⁷

Glover's approach seems to satisfy Lomasky's identifiable individual requirement, it therefore seems that a potential person can in some sense be an identifiable subject. This is, however, an error because potential persons are not subjects; that is, the person (a potential person) cannot be affected by actions which will constitute the person's conditions of existence. When we consider how these changes would be brought about we can see the claims above do not work. If we cannot change the 'health' of an embryo, then the potential person can only exist in an 'unhealthy' condition. If we can change an embryo to be healthy, then it would change the constituting conditions of the person being created and thus would not be affecting the person because we are creating a new person, making the 'identifiable' individual no longer identifiable. Thus, 'we can affirm that people cannot logically have a right to any genetic endowment, if that constitutes their identity'.²⁸ Health, among many other (and possibly all) attributes, changes the conditions constituting the potential person, such that each change is a fresh act of creation.

It could be claimed that embryos do have interests which only crystallize once they become a person. Yet this makes no logical sense for the same reason that both Glover and Glannon's 'interest' and Lomasky's 'good-for' approaches cannot work, because each approach requires a subject who benefits from the good or interest. However, the persons who supposedly possess these interests or for whom the goods are good-for are constituted by such decisions; that is, the conditions for being benefited or harmed require a person, a subject to exist, in order to determine value to that person. Just as there is no basis for saying that being brought into existence is a benefit, because no

23. Lomasky, 'Community', p. 161.

24. W. Glannon, 'Genes, Embryos and Future People', *Bioethics* 12 (2002), pp. 187–211, 193.

25. J. Glover, *Choosing Children: Genes, Disability and Design* (Oxford: Oxford University Press, 2008), p. 42.

26. Glover, 'Design', p. 42.

27. Glover, 'Design', p. 43.

28. Heyd, 'Genethics', p. 172.

subject exists who is waiting to be alive, it cannot be claimed that a potential person is benefited by being born in a particular condition. And ‘for the same reasons, we cannot hold the child is . . . an object of maleficence’²⁹; in other words, a potential person who becomes an actual person cannot be benefited or harmed by decisions which create and cause the person to exist. There are no exceptions to the statement that potential persons cannot be harmed by being caused to exist; only actual persons have interests which can be harmed and thus any regulation has to be based, and can only be based, upon the interests of extant persons.

The importance of reproductive torts

We can now turn to the changes and impact that reproductive torts would have to undergo in order to conform to the idea that potential persons have no claim over actual persons. Allowing parents (actually existing persons) to claim creates no difficulty at all because they have interests that can be affected by the actions which lead to the existence of offspring. On the other hand, allowing children who result from actions which create them to claim against someone for acts or omissions causing their existence causes a great deal of difficulty. Reproductive torts would need to be reformed because it is logically impossible for a child to be harmed or benefited by actions which create them. Furthermore, as we have seen, the necessary relationship between the person who causes the harm and the person who suffers the harm is absent, because both parties must be in a ‘single [correlative] relationship’.³⁰ It is thus unjust for medical providers to be held as having harmed someone when this cannot be the case. Moreover, if it is the action which created the individual that is harmful, then this applies regardless of whose action it is; thus, if parents decide not to abort or to prevent conception (where they have had accurate medical advice) then they are as much to blame as medical providers and should be liable under reproductive torts. As we will see when considering the US case of *Curlender v. Bio-Science Laboratories*,³¹ the logical permissibility of a child suing their parents has been raised in court before. As Rosamund Scott has highlighted, however, allowing offspring to claim against their parents would be problematic, not only because of the intra-family conflict that might arise, but because it will also ‘come up against a woman’s very personal moral interests in self-determination’.³² This of course is not a bar to the possibility of allowing offspring to claim against their parents but explains why we should avoid it.

Even with these concerns put to one side, reproductive torts still face great difficulty in terms of agreeing a basis for calculating damages. As mentioned at the beginning of this article, the fact that no one can be better off than the child who actually exists now means that the calculus of damages is impossible to determine. There is no *better or worse* situation for the child bringing the claim, because ‘nonexistence is not a state that

29. Heyd, ‘*Genethics*’, p. 109.

30. Weinrib, ‘*Formalism*’, p. 588.

31. *Curlender v Bio-sciences Laboratories* [1980] 165 Cal. Rptr. 477.

32. R. Scott, ‘Reconsidering “Wrongful Life” in England After Thirty Years: Legislative Mistakes and Unjustifiable Anomalies’, *Cambridge Law Journal* 72(1) (2013), pp. 115, 148.

can be given a value'.³³ This means that determining the condition the child would have been in (and thus being able to determine the damage caused) is not possible. It is clear that the law needs to be changed to end these confusions and contradictions, in particular the injustice of assigning liability to some parties involved in the creation of people (the medical providers) while exempting others from fault (the parents) and for creating liability for a harm that does not occur.

Further problems in English law arise from the CDCLA granting *locus standi* to disabled children to sue the medical provider who caused the disability. Some have claimed that this is discrimination against people with disabilities as it treats disability as harm.³⁴ Now obviously it cannot be discrimination because the foetus that would be aborted is a potential person and so cannot be discriminated against; that is, there is no subject of discrimination. But what the approach mooted here does make clear is that the existence of a disability *cannot* be treated any differently from a 'normal' foetus. In other words, disability cannot be conceived of as harming the potential person. Moreover, by adopting these changes, reproductive torts would be much more certain and clearly identify what the harm is, to whom it occurred, and how to remedy or compensate someone for the harm caused. It would also remove the need for the courts to answer questions relating to the value of beings that are brought into existence.

The fundamental problems of comparing existence against non-existence and the impossibility of harming a potential person remain in all cases where a child is allowed to claim. We can now discuss and analyse these reproductive torts in turn, and through this process it will be shown that the considerations that are outlined here apply across jurisdictions, because it is the very nature of allowing children to claim under reproductive torts that gives rise to the problems plaguing the courts and the injustice of creating liability when harm cannot occur.

Wrongful birth

Wrongful birth claims operate on the basis that actual persons, the parents, sought some medical intervention which was carried out incorrectly and thus were denied the option to abort or prevent the conception of a child. The parents therefore seek damages from the medical provider at fault for the harm they have suffered. The only divergence from this definition in English law is that the CDCLA allows offspring to claim for a wrongful birth as well as the parents in some circumstances.³⁵

In the English legal system, the concept of wrongful birth is embodied in case law and legislation³⁶ and includes liability for procedures during infertility treatment. Wrongful conception claims in this jurisdiction are subsumed under the title of wrongful birth cases, thus wrongful conception cases will hereafter be included in the term wrongful

33. Heyd, 'Genethics', p. 30.

34. L. Davis, *The Disability Studies Reader* (London: Routledge, 2010).

35. CDCLA 1976s.1A.

36. CDCLA 1976s.1; *MacFarlane v Tayside Health Board* [2000] 2 AC 59; *Whitehouse v Jordan* [1981] WLR 246; *Parkinson v St James and Seacroft University Hospital* [2001] EWCA Civ 530.

birth. However, the CDCLA only deals with situations where something affects the mother or offspring such as physical assault, inaccurate medical advice or a failure to diagnose a disability causing medical condition; basically this covers situations where offspring are 'born with disabilities which would not otherwise have been present'.³⁷ Thus the event, in order to qualify under the CDCLA, has to be something that inflicts the disability upon what would otherwise be healthy offspring.³⁸ Currently, it permits offspring to undertake legal action themselves against the medical provider.

One of the most prominent wrongful birth cases (although it would be known as a wrongful conception case elsewhere) is *McFarlane v. Tayside Health Board*.³⁹ This case established that parents may launch a wrongful birth claim when they would not have become pregnant but for negligence on the part of the medical organization. In *McFarlane*, the husband underwent a vasectomy which was unsuccessful and as a result his wife became pregnant. The House of Lords (now the Supreme Court) held that while it would be unjust to have the costs of healthy offspring compensated, it was fair to have the health authority pay for the damage and distress of an unexpected pregnancy. In a later wrongful birth case, *Parkinson v. St James and Seacroft University Hospital NHS Trust*,⁴⁰ the Court of Appeal held that a couple could recover the cost of an unwanted pregnancy due to the failure of a sterilization procedure *and* the cost of special care due to the disabilities of their offspring. However, the court refused to extend the coverage of *McFarlane* to cover general costs of raising a normal child, therefore restricting compensation to distress and disability. Thus, parents can claim for the costs and distress of a pregnancy and the costs of a disability if they are caused by a fault of the medical provider. This is because their interests, rights and goals have been damaged and frustrated by the wrongful action. Although, of course, strictly speaking the normal costs of raising the child should be included, they are excluded on the basis that the birth of a healthy child should not be recoverable because 'society itself must regard the balance as beneficial. It would be repugnant to its own sense of values to do otherwise. It is morally offensive to regard a normal, healthy baby as more trouble and expense than it is worth'.⁴¹ This rejection is thus based upon the interests of actually existing persons in promoting and valuing the production of children; it is not based on any supposed interests that potential people may have.

This approach has also been taken in some parts of the United States where 'in virtually all cases, courts have awarded the plaintiff mothers ... medical expenses and emotional distress damages'⁴² and have rejected damages for the 'cost of raising the unexpected child to adulthood'.⁴³ However, claims by disabled offspring, like those under the CDCLA, are dealt with differently. In the New Jersey case of *Gleitman v.*

37. CDCLA s.1(2)(b).

38. CDCLA s.1(2).

39. *MacFarlane v Tayside Health Board* [2000] 2 AC 59.

40. *Parkinson v St James and Seacroft University Hospital* [2001] EWCA Civ 530.

41. *MacFarlane v Tayside Health Board* [2000] 2 AC 59, per Lord Millett, p. 114.

42. W. Hensel, 'The Disabling Impact of Wrongful Birth and Wrongful Life Actions', *Harvard Civil Rights-Civil Liberties Law Review* 40 (2005), pp. 141–195, 151.

43. Hensel, 'Actions', p. 151.

Cosgrove,⁴⁴ a child was born with defects after the mother contracted rubella and received incorrect medical advice that rubella did not pose a risk to the foetus.⁴⁵ When the child was born the parents brought a claim of wrongful birth, matching the situation in the United Kingdom, but in addition the child brought a claim of wrongful life not of wrongful birth. The court in *Gleitman* permitted a direct claim by the child that more closely matches wrongful life cases in the United Kingdom.⁴⁶ The case of wrongful life will be considered in the subsequent section, but the divergence should be noted.

There is also the issue regarding the extent to which parents are entitled to information regarding their pregnancy. The relationship between information access about an embryo's genetic condition is discussed by Rosamund Scott in her analysis of the relationship between abortion and wrongful birth.⁴⁷ That is, in England and Wales, only serious conditions can justify abortion⁴⁸; so according to Scott, medical providers only have to convey information when it relates to a serious condition. For example, if parents would abort on the basis of their embryo having a serious medical condition, and the medical provider failed to give them that information, these facts would disclose a claim for wrongful birth.⁴⁹ If, however, the information related to something trivial, say webbed toes, then the parents cannot claim that they would have aborted on that basis and that they should have been informed.

Under the psychological personhood approach suggested here, the embryo would have no standing and so the parents could terminate whenever they wished for whatever reason, even trivial ones. Thus, the adoption of psychological personhood would also necessitate a change in abortion law to complete permissibility. Currently, the decision of what satisfies the criterion of seriousness is 'based upon the profession's assessment of the incidence of the risk and its seriousness',⁵⁰ but this would no longer be the case because the seriousness of the condition would be irrelevant. There is a strong link between the information provided and the exercising of the right (in the United States) or opportunity (in England and Wales) to abort; or failing that opportunity suing the medical provider for a wrongful birth. Thus, under the approach proposed here the medical provider would have no justification for withholding information because it is only the interests of the parents that matter; the embryo has no claims to assert against the parents. Thus, withholding information that the parents would have aborted on, as the parents are no longer constrained by the 'seriousness' criterion, could lead to a wrongful birth suit. This would greatly increase the scope and power of parental reproductive autonomy and change the law of and relating to abortion.⁵¹

44. *Gleitman v Cosgrove* [1967] 49 N.J. 22.

45. Hensel, 'Actions', p. 155.

46. Hensel, 'Actions', p. 155.

47. R. Scott, 'Prenatal Screening, Autonomy and Reasons: The Relationship Between the Law of Abortion and Wrongful Birth', *Medical Law Review* 11 (2003), p. 265.

48. Abortion Act 1967s.1(1)(d).

49. Scott, 'Birth', pp. 365–325.

50. Scott, 'Birth', p. 298.

51. Abortion will not however be discussed further in this article but this digression demonstrates the wider implications of psychological personhood.

As stated earlier, wrongful conception/birth claims are based on the harms that are clearly suffered by actual persons to their interests and would therefore not need to be changed. The basis of wrongful conception/birth claims is coherent and clearly assigns harm and damage both theoretically and legally. However, the exception which allows offspring to sue under the CDCLA⁵² must be abolished because it violates the principle that potential persons cannot be harmed by actions that should be responsible for constituting their existence. Moreover, the parties involved should be in a direct, connected single legal relationship⁵³ but the child does not exist at the time the ‘harm’ is caused and so cannot be in such a relationship. It would also prevent parents from being liable to their offspring, and explain why the current refusal to recognize such liability makes sense because it is only *parental interests* that count. No claims of harm can be recognized in law because persons exist only as a result of those actions. For the most part, the law of wrongful conception/birth would remain as it is but on the sole basis of harms to the interests of the parents (actual persons) and without the aberration of the CDCLA.

Wrongful life

Wrongful life claims are made on the basis that the child who exists does so only because the woman was denied a termination through the fault of the medical provider and that the child’s life is so full of suffering that it should not exist. Unlike wrongful conception/birth cases, wrongful life cases are brought by the child himself/herself not the child’s parents. The essential claim of wrongful life is that the offspring would have been better off not existing. In such cases, the suit for compensation is brought in the name of a (typically) severely disabled child who claims that ‘but for’ the negligence of the medical provider, the child would not have been born at all. This creates a conceptual problem of comparing something with the absence of everything. As Steininger⁵⁴ points out, wrongful life claims try to assess the damage inflicted upon offspring from living but do not have an alternative which would be worse as the courts are not prepared to compare existence to non-existence.⁵⁵ Even if the courts were prepared to make such a comparison, it is not clear how such a comparison could be made.

The case of *McKay v. Essex Area Health Authority*⁵⁶ made it clear that under English law no offspring can sue for wrongful life. Stephenson stated that it is ‘contrary to public policy, which is to preserve human life, to give a child a right not to be born except as whole, functional being’ and that it is ‘impossible to measure the damages for being born with defects’.⁵⁷ Ackner also remarked that ‘there are . . . [no damages] in any accepted sense’.⁵⁸ More forcefully, Griffiths stated that ‘I have come to the firm conclusion that

52. CDCLA 1976s.1A.

53. Weinrib, ‘Formalism’, p. 588.

54. B. Steininger, ‘Wrongful Birth and Wrongful Life: Basic Questions’, *Journal of European Tort Law* 1 (2010), p. 125.

55. Steininger, ‘Questions’, p. 152.

56. *McKay v Essex Area Health Authority* [1982] QB 1166.

57. *McKay* [1982], p. 1184.

58. *McKay* [1982], p. 1189.

our law cannot recognise a claim for “wrongful life”⁵⁹ for the ‘most compelling reason . . . is the intolerable and insoluble problem it would create in the assessment of damages’.⁶⁰ Thus, wrongful life claims have been excluded from English law for two main reasons: firstly that it is against public interest to compare existence with non-existence, and secondly that no damages can be calculated because of the impossibility of comparing loss against not existing. As Ackner commented, ‘no comparison is possible and therefore no damage can be established which a court could recognise’.⁶¹

McKay was decided before the introduction of the CDCLA, but even if that Act had been in force, such a claim would still have been rejected. As Brazier and Cave put it, ‘[w]here the essence of the claim is that the child should never have been born at all, it lies outside the scope of section 1(2)(b)’ of the CDCLA.⁶² More than simply being outside the technical remit of the Act, the Court of Appeal ruled out wrongful life claims because they were against public policy, being a violation of the sanctity of life. Although the sanctity of life is not relevant here, because actions creating potential persons create the life that would be protected by sanctity, accepting life as a potential damage would necessitate ‘an existence/nonexistence’ comparison.⁶³ This would be impossible because the court, or indeed anyone, is not capable of determining damages from an assessment which involves comparing existence and non-existence, thus no cause of action was possible.⁶⁴ This view is matched almost exactly in Germany where the Federal Constitutional Court (*Bundesverfassungsgericht*) stated that ‘the child’s existence cannot, in law, be classified as damage’.⁶⁵

However, this strict logical approach has not always been taken. Rosamund Scott, whilst accepting the logical argument regarding existence/non-existence, ‘that there was no-one or no being who could experience non-existence’,⁶⁶ argues that this framing of wrongful life claims misses the point. Scott suggests that the issue of wrongful life claims ‘lies in *being born under certain conditions*, namely when the burdens of life . . . are so severe that they outweigh any compensating goods’.⁶⁷ This view is similar to the one held by Kirby who believes that it is the suffering in wrongful life cases that should be given prominence, even going so far as to argue that ‘a life of severe and unremitting suffering is worse than non-existence’.⁶⁸ In *Curlender v. Bio-sciences Laboratories*, Jefferson claimed that the claimant ‘both exists and suffers, due to the negligence of others’.⁶⁹

59. *McKay* [1982], p. 1190.

60. *McKay* [1982], p. 1192.

61. *McKay* [1982], p. 1189.

62. M. Brazier and E. Cave, *Medicine, Patients and the Law* (5th ed., London: Penguin, 2011), p. 321.

63. N. Prialux, ‘Conceptualising Harm in the Case of the Unwanted Child’, *European Journal of Health Law* 9 (2002), pp. 337–359.

64. *McKay v Essex Area Health Authority* [1982] QB 1166.

65. *Schwangerschaftsabbruch II* [1993] BVerfGE 88, 203.

66. Scott, ‘Anomalies’, p. 129.

67. Scott, ‘Anomalies’, p. 130.

68. *Harriton v Stephens* [2006] HCA 15 at [105].

69. *Curlender v Bio-sciences Laboratories* [1980] 165 Cal. Rptr. 477, p. 488.

In the French case of *Perruche*, we can see this ‘suffering-orientated’ approach at work. The child was allowed to ‘sue his mother’s doctors because they failed to diagnose his condition while she was pregnant, [thus] denying her the choice of an abortion’.⁷⁰ The Cour de Cassation stated that ‘the issue is not his [the offspring’s] birth but his disabilities’.⁷¹ The subsequent public outrage at this decision resulted in the introduction of a new provision in the *code de l’action sociale et des familles* section L114-5, which prohibits offspring bringing a claim where they would not otherwise exist and consequently only parents can now bring a claim under French law.⁷² We can thus see that the idea that life itself could be considered ‘damage’ and thus be a cause of action in law was considered unacceptable. Scott’s rejoinder would be that it is the suffering not the life that is compensated, however, this author would still argue that this would fall foul of the ‘actions constituting existence’ problem.

In contrast, in the state of California there was a drift, albeit haphazardly, towards recognition of a child’s standing to sue. Wendy Hensel⁷³ provides a clear breakdown of the Californian court’s decisions and reasoning regarding wrongful life. One of the earliest US cases was the previously mentioned case of *Gleitman v. Cosgrove* which found that because it was ‘logically impossible’ to ‘measure the difference between life with defects against the utter void of non-existence’ wrongful life claims could not be recognized,⁷⁴ thus reflecting the position most commonly held in judicial systems. This all changed, however, in the Californian case of *Curlender v. Bio-Science Laboratories* where the Court of Appeal of California allowed offspring to claim ‘damages for the pain and suffering to be endured . . . and any special pecuniary loss resulting from the plaintiff’s condition’.⁷⁵ Interestingly, the court felt that there was ‘no sound public policy reason which should protect those parents from being answerable for the pain, suffering and misery which they have wrought upon their offspring’.⁷⁶ Thus, in some ways the Californian approach here is the most consistent because it allows ‘harmed’ offspring to claim against all those who brought it into existence; although the basis itself of such a claim is here rejected.

This prompted a swift response from the Californian state legislature to protect parents from such a claim.⁷⁷ Shortly thereafter the California Supreme Court in *Turpin v. Sortini* confirmed that the sanctity of life did not preclude claims of wrongful life and allowed damages for ‘significant medical and financial burden’⁷⁸ but rejected claims for

70. Priaulx, ‘Child’, pp. 339–340.

71. *Epx Perruche c/ Mutuelle d’assurance du corps sanitaire francais et al.* (Cour de Cassation) [2000] JCP 2000 II 10438.

72. Priaulx, ‘Child’, p. 340; *The New York Times*, October 19, 2001, Code de l’action sociale et des familles (amended to include section L114-5 in 2002).

73. Hensel, ‘Actions’.

74. Hensel, ‘Actions’, p. 155.

75. *Curlender v Bio-sciences Laboratories* [1980] 165 Cal. Rptr. 477, p. 488.

76. *Curlender* [1980], p. 487.

77. Behind, No Child Left. “Act of 2001, 20 USCA § 6301 et seq.” West 2003 (2003). The Civil Code of the State of California, California (1873) s.43.6.

78. *Turpin v Sortini* [1982] 31 Cal. 3d 220, p. 238.

general damages relating to the normal costs of existence. The claim for general damages was rejected 'because the plaintiff never had a chance of being born without her affliction, and it would be impossible to ascertain the extent of an injury in this context'.⁷⁹ However, in most other cases the 'courts [in the US] have consistently rejected wrongful life actions', reasoning that 'life burdened with defects is better than no life at all and . . . that the . . . child suffered no legally cognisable injury in being born' thus 'damages are incalculable'.⁸⁰ The exceptions (*Curlender* and *Turpin*) in California, however, 'have concluded instead that life is not always preferable to non-existence'⁸¹ or that damages are incalculable.⁸²

Rosamund Scott reaches a similar conclusion, after reviewing the arguments set out by Kirby in *Harriton v. Stephens*, Jefferson in *Curlender v. Bio-sciences Laboratories* and the combined judgment in *Perruche*, regarding how life should be treated in these situations. Scott claims that 'while it may be helpful to emphasise suffering rather than life *per se* . . . care has to be taken that the condition . . . be so severe that life will be of sub-zero quality'.⁸³ Thus, Scott's suggestion is that severity is taken as the basis of wrongful life claims and when a severe enough condition occurs, then a wrongful life claim can be brought against the medical provider. Furthermore, Scott argues that the deployment of the existential threshold in these cases is a legal construction providing a barrier against wrongful life claims. She suggests that 'emphasising that non-existence cannot be known is ultimately unhelpful, or besides the point . . . and loses sight of . . . the normative role of the construct of non-existence'.⁸⁴ Thus, Scott is essentially arguing that justice has lost out to logic.

However, Scott's suggestion is rejected by this author for two reasons. Firstly, the issue of existence/non-existence is crucial because it is this threshold that limits legal actions to those incidents which are cognizable. If we were to accept that non-existence was not a boundary to legal action, then we may be placed in the invidious position of having to take into consideration potential liability owed to people who *may never exist*. Thus, the legal prohibition on potential persons from being able to claim, in this case in reproductive torts, should stand. Moreover, the importance of the logical existential distinction should not be underestimated, as it relates to whether harms have occurred at all and to the existence of a correlative relationship between 'harmer' and 'harmed'.

This leads directly into the second objection to Scott's proposal. If we accept Scott's suggestion, and allow offspring to claim for actions that occur before their existence and use a severity threshold to avoid open-ended exposure to liability just mentioned, this would mean that wrongful life claims are permitting legal action where no harm to the offspring has occurred. It would thus be fundamentally unjust to allow medical providers

79. Hensel, 'Actions', p. 160.

80. Hensel, 'Actions', p. 161.

81. Hensel, 'Actions', p. 162.

82. Civil Code, California (1873) s.43.6.

83. Scott, 'Anomalies', p. 133.

84. Scott, 'Anomalies', p. 142.

to be liable for actions which create the particular person who is claiming against them. Medical providers would remain liable to the parents who they failed to treat with the requisite standard of care, but they would not (and should not) be liable to the offspring for creating them. Under the theory of psychological personhood set out here, the action which creates a particular person, that is the act that creates a factual as opposed to merely potential person, cannot be a harmful act to the person created. Thus, cases where wrongful life claims have been recognized should not form the basis of law and should be rejected because they rely on the false premise that life is comparable to non-existence and that those actions which constitute the existence of a person can be classified as beneficial or harmful.

Additionally, the struggle with wrongful life claims, and in some cases recognizing them, is partly down to financial considerations and distributive justice concerns. If there is no liability with the medical provider for disabled children, then the costs of their care cannot be taken from the medical provider but must come from another source. The possibility of suing their parents may provide another source of financial security, but practical considerations regarding the payment of damages may prevent claims against parents as they are likely to already be bearing the costs (or some of the costs) (but it is theoretically possible as demonstrated by Californian case of *Curlender v. Bio-Sciences Laboratories*). Moreover, as Scott mentions, parental liability would directly conflict with the mother's 'self-determination and bodily integrity'⁸⁵ therefore parental liability should be opposed for this reason as well. It thus seems (in the United States at least) that the acceptance of claims was aimed at ensuring some financial provision was in place for the care of the disabled child; this is comparable to the French court allowing the child to sue directly in *Perruche*.

However, this article has concluded that actions which create people cannot be a cause of legal action because being caused to exist cannot be a harm or damage, because it is impossible for someone to be harmed by factors constitutive of their existence and identity. Once the child is an actual person, and thus the necessary preconditions for the child to have interests is met, society (i.e. the actual persons who make up the social and political entity in which the disabled child lives) may have an obligation to the child, as an actual person. Thus, society might make provision for disabled persons and this can be achieved through general taxation, a welfare system, public health care and so on. The child does not have to sue medical providers who cause the child's existence in order to be financially supported, and indeed it has been argued here that children should not and cannot sue those who cause them to exist.

Conclusion

This article has analysed the basis of reproductive torts (wrongful conception/birth and wrongful life) through the application of the potential/actual person distinction. The law in the United Kingdom, and in other jurisdictions, on wrongful conception/birth are primarily based upon the interests of the parents who have suffered some loss or harm. This is the only sound basis for these torts. Difficulties and problems arise with wrongful life

85. Scott, 'Anomalies', p. 148.

claims and when wrongful conception/birth claims stray into being claims of the child (as under the CDCLA). Claims by the child try to bridge the potential/actual divide and thus allow claims against the medical provider for actions which constitute the child's very existence. This has been shown to be logically impossible as the person-affecting nature of torts requires that a subject exists who could have been in a better condition. However, when we are discussing actions that cause the existence of a person, then the absence of that action results in the person's *non-existence*, not the person's existence in a better state. This argument is clearly articulated when Heyd states that non-existence is 'not a state that can be *attributed* to a subject'.⁸⁶ Consequently, the use of personal harm to potential persons as the basis of reproductive torts has been ruled out as persons cannot be harmed by actions which change who will exist. This means that no potential person has or can claim an interest or a right in either existing or existing in a particular condition.

As the better or worse condition of a subject is a requirement of reproductive torts, a child who results from the actions cannot be permitted by the law because it contradicts the internal coherence of tort claims. Coherence is important because it allows those subject to the legal system to predict what conduct is permissible and what is not. This knowledge is achieved by the consistent application of the *justification* used in setting out the law. As Weinrib states the law is 'not merely a collection of posited norms or an exercise of official power but a social arrangement responsive to moral argument'.⁸⁷ Weinrib's suggests that the '[l]aw connects one person to another through the ensemble of concepts, principles, and processes that come into play when a legal claim is asserted'.⁸⁸ However, in order for the law to fulfil its function as a set of laws that can successfully be followed, these 'concepts, principles and processes'⁸⁹ must be internally compatible with each other. As we have seen in the case of reproductive torts, attempts to deal with claims by or on behalf of children have been inconsistent and at times contradictory.

Such inconsistency arises because reproductive torts do not conform to the justification for their existence. That is, the crucial relationship requirement whereby 'the plaintiff and the defendant [are treated] as correlative to one another'⁹⁰ is absent. If reproductive torts are an instrument of 'corrective justice',⁹¹ then this relationship is a crucial justification for the existence of reproductive torts in the first place. More importantly, as this relationship is the justificatory basis of reproductive tort claims, then it should inform the structure of tort claims.

When this relationship is not present, then reproductive tort claims cannot be sustained, and in the case of potential persons this requirement is never fulfilled. Thus, the structures of reproductive torts are an ad hoc construction rather than a sound, coherent and, above all, justified systematic claim. The changes to reproductive torts set out here

86. Heyd, '*Genethics*', p. 30.

87. Weinrib, '*Formalism*', p. 583.

88. Weinrib, '*Formalism*', p. 584.

89. Weinrib, '*Formalism*', p. 584.

90. Weinrib, '*Formalism*', p. 593.

91. Weinrib, '*Formalism*', p. 593.

are thus a response to this incoherence, based upon the consistent application of the justificatory theory of persons aimed at forging a ‘harmonious interrelationship among the constituents’⁹² of reproductive torts.

The current approach of tort law in judging harm and damage through comparisons between the conditions the person would have been in and the condition the person actually is in, is correct because it requires that an actual person exists who is affected; thus, tort law is person-affecting in its approach. However, allowing children to claim under any reproductive torts must logically be prohibited. In the case of the English law, the current rejection of wrongful life claims must be maintained but the CDCLA must be repealed or reformed to prevent children claiming under the wrongful birth tort. Parents must be able to claim on the basis of negligence or misconduct on the part of medical professionals because they *are* harmed; their interests and rights at the time are affected and harmed. They (as actual persons) could have been in a better situation than the one they find themselves in. Thus, parents and the harm they have suffered can be, and are, the correct basis of reproductive tort claims.

In conclusion, this article advocates a modest legal reform, namely amending the CDCLA to prevent claims by children when actions cause disabilities, but it provides arguments both to engender this reform and to maintain the rejection of any claims by children under reproductive torts (and under wrongful life in particular). The implications of the arguments presented here, in relation to both the and ethical basis of claims, are far reaching in other reproductive issues (such as genetic modification, abortion and procreative decisions) as in all cases the potential person cannot be the basis of any claims or restrictions on the conduct of the prospective parents. Only the effects on other actual persons or the conduct of the parents themselves (e.g. on the basis of virtue ethics or a standard of behaviour) can ground legal sanction, particularly as these different individuals can form correlative relationships with each other that potential and actual persons cannot. These wider implications are beyond the scope of this article and thus would need to be (and deserve to be) considered in separate articles. For the moment, this article concludes with the desirability of reforming the CDCLA suggested earlier and the rejection of any claims by children for actions that are responsible for their existence, because potential persons cannot be affected by actions bringing them to existence. Basically it comes down to this; there is no justification for creating liability to offspring for actions which constitute the offspring’s very existence.

Acknowledgements

I would like to thank Professor Søren Holm and Professor Mätti Hayry for their comments regarding this article and the reviewers for their constructive comments.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

92. Weinrib, ‘*Formalism*’, p. 593.