The London School of Economics and Political Science

By any means necessary?
A liberal theory of social justice activism

Temitayo Ogunye

A thesis submitted to the Department of Government of the London School of Economics and Political Science for the degree of Doctor of Philosophy, London, August 2021

Declaration

I certify that the thesis I have presented for examination for the MPhil/PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others.

I declare that my thesis consists of 63,033 words.

Abstract

Political philosophers have produced a wide variety of competing theories of the perfectly just society, but they have paid less attention to the question of how perfect social justice is to be achieved or worked towards. This is especially odd because most, perhaps all, societies are unjust by any plausible yardstick. Let us call any attempt to advance social justice or remedy social injustice *social justice activism*. My aim in this dissertation is to develop a theory of social justice activism. I make three contributions in particular. First, I seek to establish the grounds of the duty to remedy social injustice. I argue that we should appeal to multiple principles to ground remedial duties. Second, I argue that, in order to understand how to remedy social injustice, we must first understand the different kinds of social phenomena that can underpin social injustice. I identify three such phenomena: laws, social norms, and stereotypes. Third, I consider how activism that aims to remedy social injustice underpinned by these three mechanisms should be practised, devoting a chapter to each one. Regarding law, I explore the ethics of activism that seeks to change the law via means that are either illegal or violent, and I argue that the practice of such activism is more morally constrained than is standardly assumed. Regarding social norms, I argue that law and policy will not always be sufficient to remedy injustice caused by social norms, and so ordinary citizens will sometimes need to intervene to change unjust social norms in ways that are not mediated by the state. Finally, because the operation of stereotypes is rather subtle and mysterious, I seek to better understand exactly how it is they generate social injustice. I then draw out some implications of this investigation for the practice of social justice activism.

Acknowledgements

From the outside, writing a PhD looks like a very solitary activity, and, in many ways, it is. But it is also a task that could not completed without a lot of help. I have many people and groups to thank.

Firstly, I'm grateful to the LSE for providing the funding that made writing this PhD possible. In addition, I'm grateful to all of the security, cleaning, and professional services staff at the LSE who I've interacted with so often over the years, and who have always been kind and helpful.

I'd also like to thank the members of the Political Theory Group at the LSE. This group has provided a socially relaxed and intellectually rigorous environment in which to think, argue, present, and grow. I'll miss being part of it. In particular, I'd like to thank all of the political theory PhD students and Fellows past and present who I have shared my time at the LSE with, especially Anahi, Paola, Kaveh, Fergus, Cain, Vincent, David, Signy, Will, Hwa, and Luke. They have been a regular source of advice, ideas, critique, and fun. Many of them have become friends. I'm especially grateful to Antoine Louette. We started the PhD together, struck up a bond immediately, and have been friends throughout the journey. I hope we remain friends for much, much longer.

I owe very special thanks to my supervisor, Laura Valentini. For the past few years, Laura has always been on hand to provide guidance, encouragement, inspiration, and challenge. She has remained patient with me as I've taken many detours and entered many intellectual cul-de-sacs, and she has retained faith in me during those periods when I lacked faith in myself. This thesis would not exist without her. I'm also very grateful to my PhD adviser, Kai Spiekermann, who has always been extremely generous with his time and wisdom.

Most importantly, I want to thank those who are nearest and dearest to me. To my partner, Laura: thank you for your love, support, and faith, for bearing with me when work wasn't so going well, and for always having my back. To my mum: thank you for giving me life, for raising me in challenging circumstances, and for your unconditional love.

Table of Contents

Introduction

> ideal thinkers who want to have some impact on reality should pay more attention to issues of transition.
>
> Wolff 1998:113

Justice is probably the most commonly invoked value in contemporary political philosophy, and social justice is perhaps the most frequently discussed variant of this value. Rival theories of social justice abound, varying across multiple dimensions, but one feature very many accounts share is that they focus on describing and defending what each account takes to be the perfectly just society. Borrowing terminology coined by John Rawls, let us call any theory that purports to describe the perfectly just society an *ideal theory* of social justice, and let us call any account that seeks to explain how it is that perfect social justice is to be achieved or worked towards a *nonideal theory*.

Rawls famously focuses most of his scholarly attention on developing the former, and only a very small proportion of his work is concerned with the latter. This is not because Rawls holds the implausible view that nonideal theory is somehow uninteresting or unimportant. He insists that the issues covered by nonideal theory are "the pressing and urgent matters… the things we are faced with in everyday life". Instead, Rawls's stated rationale for focusing primarily on ideal theory over nonideal theory is that this "offers the only basis for the systematic grasp of these more pressing problems" (Rawls 1999a:8). As he puts it: "until the ideal is identified… nonideal theory lacks an objective, an aim, by reference to which its queries can be answered" (Rawls 1999b:90). Perhaps unsurprisingly given his influence in contemporary political philosophy, Rawls's position on the priority of ideal over nonideal theory is emblematic of a wider literature in which there are very many rival accounts of the perfectly just society, but comparatively little has been said about how it is that perfect social justice is to be achieved or progressed towards from real-world circumstances that all agree are unjust.

My aim in this dissertation is to develop an account of how to advance social justice and remedy social injustice. Let us label any attempt to advance social justice and remedy social injustice *social justice activism*. Social justice activism is the subject of nonideal theory and so the contribution I intend to make in this dissertation can be described either as an account of nonideal theory or as a theory of social justice activism;

I will typically use the latter formulation. More specifically, my aim in this dissertation is to buck the trend in the literature described in the previous paragraph in two ways: not only will I undertake the neglected task of developing a theory of social justice activism, but I will do so without first endorsing a fully-developed theory of the perfectly just society. In this introduction, I will first provide some context for the aim just described. Then, I will outline the account of social justice activism I will develop to deliver on that aim.

0.1. The challenge of developing of theory of social justice activism without endorsing a fully-developed ideal theory

To understand the context for my aim of developing a theory of social justice activism without first endorsing a fully-developed ideal theory of social justice, we must first understand the rationale for prioritising ideal over nonideal theory in more detail. The rationale is not that one cannot tell what counts as an injustice – and so what one would wish to see remedied – in the absence of a theory of the perfectly just society. Ideal theory is not necessary to identify the most egregious and uncontroversial social injustices. As Amartya Sen observes, just as we don't need to know that Everest is the tallest mountain in the world in order to compare two smaller mountains and to see which one is taller, so we do not need to know what perfect social justice entails in order to know that a society in which slavery has been successfully abolished, say, constitutes an improvement on an otherwise identical society in which slavery is widely practised (Sen 2006:221-2). Instead, the rationale for prioritising ideal over nonideal theory is that we do not want whatever incremental steps we take in the short term to foreclose the possibility of much more ambitious gains at a later date. In other words, we need to chart a viable route – a "navigation map" (Robeyns 2012:160) – from where we are to where we eventually want to be. But, clearly, in order to do that we need to know where it is that we eventually want to be – that is, we need to know what perfect social justice looks like. As A. John Simmons notes, building on Sen's Everest analogy,

> which of two smaller 'peaks' of justice is higher (or more just) is a judgment that matters conclusively only if they are both equally feasible paths to the highest peak of perfect justice. And in order to endorse a route to that highest peak, we certainly *do* need to know which one that highest peak is (Simmons 2010:35 italics in original).

However, given the sheer variety of rival accounts of perfect social justice that political philosophers have produced, the rationale for prioritising ideal over nonideal theory just offered would seem to suggest that there is simply no prospect for agreement about a theory of social justice activism. If this is the case, then nonideal theory is likely to be inescapably partisan; each of the competing social justice factions will have a different account of the final destination to be arrived at and so of the route to get there.

0.2. Overcoming the challenge by assuming the core of ideal theory

I think that it is possible to develop a theory of social justice activism that is more ecumenical than this pessimistic but tempting thought suggests. I say this because, though there are exceptions, it seems to me that there is far more agreement amongst political philosophers regarding what perfect social justice entails than is typically acknowledged. Put simply, while there may be a very wide variety of final destinations, starting from where we are, many theories demand that – for the time being at least – we move in the same direction on the same motorway and make the same pit-stops.

There are three areas of agreement I would like to highlight in particular. First, arguably many political philosophers would agree that, in a perfectly just society, all citizens would enjoy equal and secure access to the content of certain rights. These include the right to bodily integrity, the right to freedom of thought, the right to freedom of expression, the right to freedom of association, and the right to vote and stand for political office. Second, it seems to me that many political philosophers would agree that, in a perfectly just society, all citizens would be entitled to a minimum level of resources, such that no one is forced by material necessity to starve or live a life that lacks basic decency. Of course, many theories of perfect social justice have prescriptions related to access to and the distribution of resources that are much more demanding and complicated than this. Nevertheless, I think that secure access to a minimum level of resources is the very least most ideal theories would require. Finally, amongst those political philosophers who allow that a perfectly just society would contain unequal positions when it comes to authority or income, many would agree that all citizens should

have fair opportunities to develop their talents and compete for these more advantaged positions.[1]

I want to suggest that these three requirements – that all citizens enjoy equal basic civil and political rights; that all citizens are entitled to at least the minimum level of resources necessary to avoid starvation and degradation; and that, if unequal positions of authority and income are permitted at all, all must have fair opportunities to compete for them – would be included in any plausible and complete ideal theory. Put another way, while there are many rival ideal theories of social justice which differ in various ways, all plausible and complete ideal theories will have certain things in common. I want to suggest that the three requirements just mentioned will be amongst these things held in common. Let us call those features which are common to all plausible and complete theories of the perfectly just society *the core of ideal theory*.[2] Though I provide some arguments in support of some features of the core of ideal theory in chapter 4 of this dissertation, for the most part I will seek to develop a theory of social justice activism on the assumption that the core of ideal theory is settled and so not up for debate. I will take no stand on what perfect social justice requires over and above the core of ideal theory.

Most, perhaps all, contemporary societies fall short of the core of ideal theory as I have described it. Moreover, the fact that we are in possession of the core of ideal theory means that it is possible to develop an account of social justice activism that is ecumenical, despite the ongoing debate about what perfect social justice entails. That is, we can begin to address the 'pressing and urgent' questions of nonideal theory without a fully-developed account of the ideal. This is compatible with the rationale for prioritising ideal over nonideal theory. The claim is not that we do not need ideal theory, but that "enough of ideal theory is settled that we can already begin to derive from it (conjoined

[1] The scope of political philosophers who can endorse these commitments ranges from (some) libertarians (e.g., Lomasky 1987) to liberal egalitarians (e.g., Rawls 1999a). Liberal socialists (e.g., Cohen 2009) can endorse the first two claims. However, because socialists are generally very suspicious of inequality in relation to authority and income, I have phrased the third commitment in the conditional form in the hope of describing the core of ideal theory in as ecumenical way as possible.

[2] Structurally, though not necessarily in terms of content, what I call the core of ideal theory is very similar to what Gopal Sreenivasan calls "non-ideal theory as *anticipatory* theory". According to Sreenivasan, the latter makes "assumptions about the minimum requirements that *any plausible and complete* ideal theory of justice will include" (Sreenivasan 2007:221 italics in original).

with our social scientific knowledge) the rudiments of nonideal theory" (Simmons 2010:36, also see Stemplowska and Swift: 2012:380).

I acknowledge that there will be those who do not endorse the core of ideal theory as I have described it, either because they disagree with the substance of the requirements or with the conceptual framework and normative vocabulary with which they have been articulated. And I accept that the theory of social justice activism I develop in this dissertation may not be able to speak to those who disagree with the three requirements I have identified. Nevertheless, given how widely endorsed these requirements are, it seems to me that a theory of social justice activism based on them will still be relatively ecumenical.

0.3. A liberal theory of social justice activism: an outline of the account

Having introduced the core of ideal theory on which my account of social justice activism will be based, let me now turn to outline the account itself. The theory of social justice activism I develop in this dissertation can be divided into three parts. First, I seek to establish the grounds of the duty to remedy social injustice. Though the question of how social injustice is to be remedied has been underexplored in contemporary political philosophy, the idea that most if not all societies are unjust and that the citizens of such societies have duties to remedy the injustice in their midst seems to be a common and perhaps even the consensus view. This raises the question: what are the grounds of these duties? The *grounds* of a duty are the values and principles that are invoked to explain or justify it. In chapter 1 of this dissertation, I argue that focusing exclusively on any one of the principles that have been advanced as grounds for remedial duties in the literature generates counterintuitive implications which can only be avoided by appealing to multiple principles. Moreover, according to the approach I advocate, and unlike some other accounts, the various values and principles that are invoked to ground remedial duties can combine and interact. This is why I call the account I defend in chapter 1 'an intersectional approach to grounding remedial duties'.

Second, in chapter 2, I argue that, in order to understand how to remedy social injustice, we must first understand the different kinds of social phenomena that can underpin injustice. I identify three such phenomena: laws, social norms, and stereotypes. We are perhaps most familiar with social injustice generated by law – examples include laws that denied women or black people the right to vote. Social norms are more informal

rules of conduct and can also underpin social injustice – consider, for instance, how gender norms can limit women's opportunities in the workplace. Finally, stereotypes can generate social injustice via the phenomenon commonly known as 'implicit bias'. I refer to these social phenomena as 'the three sites of social justice activism' because I will explore the ethics of activism that seeks to intervene in each of them in chapters 3-5. In the process of introducing these three social phenomena as potential vehicles of social injustice, I also suggest amendments to and an integration of the leading theory of social norms on the one hand and of law on the other. Thus, as well as laying the foundations for the chapters to come, an additional aim of chapter 2 is to contribute to debates in social ontology.

In the third part of the theory of social justice activism I develop in this dissertation, I explore how social justice activism that intervenes in each of the three social phenomena introduced in chapter 2 should be practised. This third part is itself further divided into three as I devote a chapter to each mechanism of social injustice. Regarding law, in chapter 3, I explore the ethics of forms of activism that seek to change the law via means that are either illegal or violent. Examples of such means include civil disobedience, rioting, militancy, and civil war. I call such forms of activism 'radical activism'. To the extent that the ethics of radical activism has been discussed in the literature, it has typically been considered in light of the fact that states claim they have the right to rule and that their citizens have a correlative duty to obey the state's commands because the state has issued them. That is, states claim to have *political authority*. This is why I focus on radical activism: its raises distinctive political authority-related concerns. In chapter 3, I highlight that the standard framing of the ethics of radical activism presupposes a binary account of political authority, argue that we should disaggregate political authority and so reject a binary account, and draw out the implications of this disaggregation for the practice of radical activism.

Regarding social norms, in chapter 4, I explore how it is that we should encourage people to abandon prevailing unjust social norms and to comply with new, more just social norms instead. I call this kind of activism 'informal activism'. My discussion of informal activism is divided into two parts. First, I highlight the limits of a widely assumed view, according to which citizens' duties to remedy social injustice must be discharged by encouraging the state to change law and policy. The fact that this view is limited suggests that citizens should sometimes seek to remedy social injustice via means that

are not mediated by the state. In the second part of my discussion in chapter 4, I explore the question of how citizens should engage in informal activism that is not mediated by the state. I argue that the value that should guide the practice of informal activism is respect for moral personhood, and highlight what this implies about the forms of influence that informal activists may engage in.

Of the three social phenomena I introduce in chapter 2, it is the operation of stereotypes that is the most subtle and mysterious. Thus, in chapter 5, I seek to better understand the role that stereotypes play in generating social injustice and to draw out the implications of this role for the practice of social justice activism. My strategy is to begin by examining a recent debate between Elizabeth Anderson and Tommie Shelby regarding the ethics of racial integration, and then to use the insights this investigation yields as a platform on which to build an account of activism that I call 'stereotype activism'. Stereotype activism involves attempting to remedy social injustice by changing the stereotypes that are present in a culture. I argue that this will often involve intervening to change the social environment that people inhabit and defend this form of activism against the charge that is it manipulative.

In chapter 6, I conclude by reviewing my argument: I explain why I call my account 'a liberal theory of social justice activism', and I draw out some implications of my view for the practice of activism in western societies.

0.4. Ambitions and limitations

As should be clear from the outline given in the previous subsection, much of my attention in this dissertation will be focused the question of how social justice activism should be engaged in and the precise form it should take in different circumstances. I think that much more can be said in response to this question than is typically assumed, and my ambition is to provide as much practical guidance as I can. That said, it is also worth emphasising that there is of course a limit to the amount of guidance that any theory designed to apply to a general class of cases can provide. Despite my best efforts, I'm sure that I will not reach that limit in this piece of work. In fact, an additional ambition of this dissertation is to invite others to further probe and explore how much more political philosophy can contribute in answer to the pressing and urgent question 'how should we promote social justice and remedy social injustice?'.

1. An intersectional approach to grounding remedial duties

> the processes through which men incur obligations are unavoidably pluralistic.
>
> Walzer 1970:15

Introduction

A common – perhaps the consensus – view in political philosophy is that most contemporary societies are unjust, and the citizens of unjust societies have duties to remedy the injustice in their midst. There are at least two questions one might ask about the suggestion that citizens have such duties:

1. *Grounds*: What values or principles explain or justify these duties? *What are their grounds*?
2. *Content*: What does discharging these duties involve? *What is their content?*

As I already mentioned, social justice activism refers to any attempt to advance social justice and remedy social injustice. Let us call duties to engage in social justice activism *remedial duties*. The purpose of this chapter is to establish the grounds of citizens' remedial duties. To do so, I examine five principles that I take to be candidate grounds for duties to engage in social justice activism: the self-regarding principle; the general, other-regarding principle; the contribution principle; the associative principle; and the principle of fairness. My primary aim is not to offer a full-blown assessment of these principles themselves. I will suggest some considerations which I think speak in favour of these principles, and I will endeavour to identify what seems to me to be the most plausible version of each of them. But I am largely sympathetic to these principles, and I assume that they are valid in some shape or form. Instead, my aim is to show that focusing exclusively on any one of these principles when grounding remedial duties generates counterintuitive implications which can only be avoided by appealing to multiple principles.

My claim is not that many scholars explicitly endorse a single-principle approach or would do so when pressed. But a pluralistic approach to grounding remedial duties is

not commonly explicitly endorsed either. Hence, one way to understand this chapter is as an effort to make explicit a view that is often left unsaid or implicit, by highlighting the implications of denying the view. I am also not the first to defend a multi-principle approach (e.g., Miller 2001, Caney 2010). However, according to the approach I advocate and unlike some other accounts, the various values and principles that are invoked to ground remedial duties can combine and interact. This is why I call the account I defend 'an intersectional approach to grounding remedial duties'.

One advantage of the intersectional approach is that it is able to progress our understanding of the duty to engage in social justice activism beyond just its grounds. Those who discuss citizens' duties to remedy social injustice often focus almost exclusively on question 1 above, the grounds question. Part of the reason for this is that the answer to question 2 – the content question – is said to be indeterminate and variable, and so impossible to specify in advance. For instance, Iris Marion Young argues that citizens of unjust societies have duties to remedy the injustice in their midst, but insists that "[n]o philosophy can tell actors just what we ought to do to discharge our responsibility" (Young 2011:124) and "[i]t is up to the agents who have a responsibility to decide what to do to discharge it within the limits of other moral considerations" (Young 2011:143). Duties that offer latitude or discretion of this kind are often called *imperfect duties*.

According to Candice Delmas, another advocate of citizens' duties to remedy social injustice, the duty she defends is "imperfect: its scope and content are indeterminate; and one carries considerable discretion regarding the actions that count as meeting the duty" (Delmas 2014:482). Similarly, Tommie Shelby argues that citizens of the United States of America (USA) have duties to remedy the unjust conditions that characterise their society, but adds that "[a]s an imperfect duty" it "allows each some discretion in how he or she will carry it out" (Shelby 2016:223). Examples of acts that might count as discharging one's remedial duties include (but are by no means limited to) voting for a political party committed to advancing social justice, donating one's time or money to an organisation dedicated to remedying social injustice, or engaging in protests and demonstrations to highlight one's opposition to social injustice.

It is true that there is only so much that can be said about the content of remedial duties in the abstract, because it will vary across situations and duty-bearers. Nevertheless, I think that much more can be said than has been to date. In relation to this,

it is worth noting that question 2 above – the content question – actually conceals an additional question within it:

3. *Weight*: What are the burdens associated with discharging the duties? *How weighty are they?*

Despite the indeterminacy and variability of answers to the content question, I think it is possible to make some progress on question 3 – the weight question – in advance of specific circumstances.[3] The intersectional approach to grounding remedial duties helps us to make this progress. According to the intersectional approach, different principles pick out different sets of individuals as duty-bearers and the same individual may be picked out as duty-bearer by more than one principle. I argue that, all other things being equal, the more principles that pick you out as a duty-bearer, the weightier your duty to engage in social justice activism. Moreover, not only does the intersectional approach assist us in answering the weight question, but it also helps us in making progress on question 2 above, the content question.

The chapter proceeds as follows. In sections 1.1-1.5, I examine the five principles which I consider to be candidates as grounds of the duty to engage in social justice activism I mentioned earlier. There are two stages to my discussion of each of these principles. First, I offer some considerations which I think speak in favour of the principle and, if necessary, suggest amendments to the principle so that it appears in its most plausible light. Second, I highlight why focusing exclusively on any one principle generates counterintuitive implications. This suggests that we should take a pluralistic or multi-principle approach to grounding remedial duties. In section 1.6, I develop and elaborate on the particular multi-principle approach I favour – the intersectional approach – by comparing it with other multi-principle approaches, and by highlighting how it helps us to make progress on the weight and content questions. Section 1.7 concludes.

[3] I also think it is possible to provide more guidance about the kinds of actions that ought to be performed in different kinds of situations. This will be the topic of chapters 3, 4, and 5.

1.1. The self-regarding principle

The first principle for grounding the duty to engage in social justice activism that I examine is what I call the *self-regarding principle*. The self-regarding principle refers to an agent's own interests or worth in order to ground remedial duties. Because this principle is self-orientated, it picks out victims of injustice as duty-bearers. The self-regarding principle seems to me to have a unique and important feature: it recognises the special relation in which victims stand in *vis-à-vis* social injustice and the significance of victims being active agents in the project of remedying the unjust conditions they face. As we shall see, at least one account of the grounds of the duty to engage in social justice activism focuses on the duties of beneficiaries of injustice. The central failing of an account focused exclusively on the duties of beneficiaries is the implication that victims (who are included amongst but do not exhaust non-beneficiaries) are merely passive recipients in the project of remedying the conditions of social injustice they face, simply moral patients – beings morality requires others to do things for – rather than moral agents (see O'Neill 1998). The self-regarding approach does not suffer from this drawback.

According to what seems to me to be the most plausible version of the self-regarding principle, all individuals ought to have an appropriate level of concern for their own objective well-being. However, because conditions of social injustice harm the well-being of the victims of social injustice, the most appropriate manifestation of victims' concern for their well-being under conditions of social injustice will sometimes be to attempt to remedy the injustice in question (Silvermint 2013). I say 'sometimes' because appropriate concern for one's well-being will not always call for victims to engage in social justice activism. The reason is that much social justice activism will be burdensome – requiring agents to invest time, effort, financial resources, or run the risk of incurring sanctions. Let's assume that there is a minimum well-being threshold below which morality cannot *require* anybody to fall. I will not attempt to specify where the minimum well-being threshold should be drawn. Nevertheless, it seems likely that, in conditions of social injustice, there will be some – perhaps many – who will be at, near, or already below this threshold, wherever it is set, and engaging in social justice activism will likely worsen their situation by pushing them (further) below it. Under these conditions, the self-regarding principle will not speak in favour of engaging in social justice activism.

When it comes to grounding remedial duties, the obvious problem with focusing exclusively on the self-regarding principle is that it has nothing to say about the duties of non-victims. I don't mean to suggest that those who argue that victims have self-regarding duties believe these to be the only duties generated by unjust circumstances. The point I am making is that the self-regarding principle cannot justify or explain these other duties. To do so, we must invoke additional principles – that is, we must take a multi-principle approach.

1.2. The general, other-regarding principle

Let us then turn to a second candidate principle for grounding duties to engage in social justice activism: what I will call the *general, other regarding principle*. The general, other regarding principle applies to all moral agents simply as such and requires that they work to secure social justice for all other moral agents. Very few people deny that something like the general, other-regarding principle exists. It captures the foundational thought that, not only are all entitled to social justice, but social justice also makes demands of us all. In fact, achieving social justice would not be possible if ordinary people did not contribute to its realisation.

There are a number of forms that the general, other-regarding principle can take. First, there is what we might call the *weak version*, according to which the general, other-regarding principle is not very demanding. For instance, John Rawls argues that all individuals are bound by the natural duty of justice to "assist in the establishment of just arrangements when they do not exist, at least when this can be done at little cost to ourselves" (Rawls 1999a:293-4). This duty applies to all irrespective of their voluntary acts, social relations, or institutional involvement (Rawls 1999a:98-99), and Rawls presents civil disobedience as one possible instantiation of the duty in conditions of social injustice (Rawls 1999a:319). The problem with this version of the general, other-regarding principle has to do with the weight of the duty that it generates. By *weight*, I mean the extent of the burdens or costs that can be associated with discharging a duty. I assume an objective understanding of cost such that there are certain experiences – physical pain, social disapproval, or a reduction in material resources, for instance – that qualify as costs for all people, regardless of any given person's subjective attitude towards these experiences. The weightier a duty is, the greater the cost an agent can be required to bear to discharge it.

Suppose we take Rawls's 'little cost' caveat to refer to some absolute level of burden that the natural duty of justice must not exceed. To see the implications of this caveat, consider the following example. Imagine a deeply religious society in which citizens interpret their faith in such a way that leads them to believe that female genital cutting (FGC) is required. The citizens are also strongly influenced by the religious clerics of their society, and the most senior cleric in particular. Suppose that the most senior cleric has come to doubt that the correct interpretation of the society's religion requires FGC and suppose that, were he to express his view, he would persuade a large proportion of his fellow citizens to change their minds and cease the practice. Finally, imagine that this senior cleric's view is seriously out of step with the rest of the clerics and that his colleagues will shun and ostracise him if he expressed his view. Clerics train, live, and work together and form an intimate and intense community; for a cleric to be ostracised by this community is more than a 'little cost'.

Given the seriousness of the social injustice that is FGC, it seems to me that this cleric has a duty to speak his mind, despite the costs involved. But, at least according to the interpretation of Rawls's view under discussion, this cleric does not have such a duty because discharging it would involve incurring more than 'little costs'. I take this to be a counterintuitive implication of Rawls's view. Hence, I agree with David Lyons, who calls Rawls's 'little cost' proviso a "wide escape clause" and notes that "one might have supposed that, the greater the injustice, the greater the sacrifice that morality might call on us to risk" (Lyons 2015:552). What this suggests is that the weak version of the general, other-regarding duty – or at least Rawls's account of it – is insufficiently weighty.

Let us therefore consider a second version of the general, other-regarding principle. According to what we might call the *strong version*, the general, other-regarding principle is very demanding. Peter Singer, for instance, holds that all people have a duty to prevent morally bad things from happening, so long as this does not involve sacrificing "anything of comparable moral importance". This duty, Singer says, takes no "account of proximity or distance" (Singer 1972:231). One implication of Singer's position, at least with respect to extreme poverty, is that the rich should give to the poor "until [they] would cause as much suffering to [themselves] or [their] dependents as [they] would

relieve by [their] gift" (Singer 1972:241).[4] As with the weak version, the problem with this version of the general, other-regarding principle has to do with the weight of the duty that it generates. More specifically, the strong version of the general, other-regarding principle seems excessively weighty. While it accommodates the plausible thought that costs or burdens should rise in line with the badness of the situation to be avoided, it does not impose an upper cost limit after which the duty to engage in social justice activism is defeated. This means that if a truly horrendous situation can be avoided by sacrifices that, though enormous for an individual agent, pale in comparison to the impending disaster, then that level of sacrifice is morally required of the individual.

One way to solve the problems associated with both the weak and the strong versions of the general, other-regarding principle would be to say that the weightiness of the duty generated by the principle should be calibrated to the seriousness of the injustice in question, but there is a cap on the level of cost an individual can be required to bear to discharge this duty.[5] Once this cost level has been surpassed, the duty is extinguished. I will not specify where this cost level should be set, but it is plausible to think that it will be in line with the minimum well-being threshold referred to earlier. What I will say is that the cost cap should be proportional, not flat-rate. This is because agents vary a great deal with respect to the impact that the same raw amount of cost will have on their lives.[6] Let us call this the *calibrated version* of the general, other-regarding principle.

---

[4] It is worth noting that Singer's focus is on the duty to eliminate extreme *global* poverty. Nevertheless, his account seems relevant and applicable to my purposes.

[5] Though Singer is himself committed to the position attributed to him in the text, he also offers a more "moderate" version of his duty. This version requires people to "prevent bad occurrences unless, to do so, we have to sacrifice something morally significant" (Singer 1972:241). Because this version of the duty does not have a comparative element, it seems consistent with the idea of a cap on the sacrifices that morality can require people to make, though of course much will depend on what Singer takes to be 'morally significant'.

[6] Former Labour MP and British Foreign Secretary David Miliband brought this out nicely in a contribution to a debate on welfare reform in the House of Commons in 2013:

> The measures before us raise £3.7 billion from poor and lower-middle-income people in 2015-16. The Chancellor cut tax relief for pension contributions by wealthier people, but by how much? It was by £200 million in 2013-14 and £600 million in 2015-16. The cumulative saving from the richest between now and 2015-16 is £1.1 billion; the cumulative saving from those on lower-middle incomes on benefits and tax credits is £5.6 billion… Taking five times as much from lower and middle-income Britain as from the richest in Britain is not equality of sacrifice. The Chancellor

Having identified what seems to me to be most plausible version of the general, other-regarding principle, let me now highlight the problem with focusing exclusively on it to ground remedial duties: it fails to account for the strong intuition that people's duties to respond to some bad state of affairs should be informed by their relationship to it. Call this *the problem of ignoring relations*. Of course, someone committed to focusing exclusively on the general, other regarding principle to ground remedial duties could bite the bullet and accept that their view is in conflict with this deeply held intuition. One striking – and, it seems to me, theoretically very costly – implication of this view is that victims of injustice who fail to discharge their duty to engage in social justice activism are guilty of the same moral failing as perpetrators or beneficiaries who do the same (Silvermint 2013:415).

It is tempting to think that we can avoid the problem of ignoring relations by fine-tuning the general, other-regarding principle so as to make it more sensitive to duty-bearers' relationships to injustice. Andrea Sangiovanni, for instance, argues that the "[natural] duty increases in strength – and hence in its capacity to outweigh or override competing considerations – the more we benefit from injustice and the more we are complicit in the preservation of the current (unjust) institutions" (Sangiovanni 2015:353). The problem with this suggestion is it fails to remain faithful to the natural duty of justice, which, recall, binds all moral agents simply as such, irrespective of voluntary acts or institutional involvement. Instead, Sangiovanni has invoked two additional principles – benefitting from injustice and complicity in injustice – to solve the problem of ignoring relations. I don't object to combining multiple principles in this way – I advocate such an approach myself – I just think that it is better to acknowledge the fact that we are doing so.

1.3. The contribution principle

The third principle I will examine is what I will call the *contribution principle*. It holds that those who count as contributors to social injustice have a duty to remedy it. There are at

---

reminds me of the man at the top of a ladder in a 1929 election poster. The man at the bottom of the ladder has got water up to his neck, and the man at the top shouts, "Equality of sacrifice—let's all go down one rung!" It is not equality of sacrifice when you are up to your neck in water (D. Miliband in House of Commons 2013:217).

least two forms that the contribution principle can take. The first is what we might call the *weak contribution principle*. On this view, to count as a contributor, one need not intend that the unjust situation come about, nor must it be the case that the situation would not have occurred but for one's contribution. Participation is all that is necessary. Iris Marion Young grounds the duty to engage in social justice activism in the weak contribution principle. She identifies what she takes to be a distinctive class of injustices which cannot be attributed to the wrongful actions of a limited set of specific agents, but is instead the result of large-scale social processes produced by the actions of very many individuals. Young argues that all individuals who participate in these processes have a shared duty to transform them so that they no longer have unjust results (Young 2011).

It seems to me that the main problem with the weak contribution principle relates to the very low threshold the principle sets for one to count as a contributor. Why, one might ask, should mere participation be thought sufficient to ground remedial duties on contribution grounds? Insofar as there is a considered judgement underpinning the contribution principle, it seems to be that people should bear the burdens (or reap the rewards) associated with situations to the extent that their actions can be credited with bringing about those states of affairs, not that people should bear the burdens (or reap the rewards) associated with situations when they participate in the large-scale social process that bring those situations about. Intuitions in support of the latter view seem to be much less widely shared or deeply held than intuitions in support of the former view.

One possible answer to the question just posed about why meeting such a low threshold should ground remedial duties is that sometimes there are no agents in particular who can be credited with bringing about an unjust state of affairs, or those who can be have left the scene, and so we need the weak contribution principle so as to avoid the counterintuitive implication that nobody has any duties to remedy the injustice in such situations. But, in response, we can say that the general, other-regarding principle already forestalls that counterintuitive implication. In fact, given that all or almost all citizens in a modern society will participate in the large-scale social processes that produce social injustice in that society, it seems that the weak contribution principle is in effect indistinguishable from the general, other-regarding principle.

In light of this problem with the weak contribution principle, it seems to me that a second form of the contribution principle is more plausible. This version of the principle sets a more demanding threshold for someone to count as a contributor and, for that

reason, I will call it the *strong contribution principle*. According to this view, to count as a contributor, a person must have engaged in an act (or omission) that was under their control, which foreseeably resulted in the occurrence of social injustice, and for which they can be blamed.[7] And, where more than one citizen meets the threshold to count as a contributor, the weight of each citizen's remedial duty is determined by the extent of their contribution.

To illustrate the difference between the weak and the strong contribution principles, consider the following example. Imagine a society that is unjust because the major social institutions of the society – such as the labour or housing markets, or the criminal justice or education systems – are set up in such a way that they foreseeably produce a situation in which some citizens have far less secure access to their civil and political rights, are more likely to be in poverty, and tend to lack fair opportunities to compete for more advantaged positions. Many ordinary citizens of this society disapprove of the injustice in their midst, but all citizens inevitably and unavoidably participate in the social processes that produce and sustain it. They participate in the housing and labour markets, for instance. The regime governing this society is not very responsive to the people and the members of the governing class are drawn exclusively from an elite that is at best indifferent to and at worst approving of the injustice that characterises the society. The government could change the laws and policies of the society so as to bring the injustice to an end, but it simply chooses not to.

According to the weak version of the contribution principle, the fact that all citizens of the society participate in the large-scale social processes which produce the social injustice means that all citizens have contribution-grounded duties to remedy the injustice in their midst. The strong contribution principle does not yield this judgement. Instead, it says that only those members of the government who enact the laws and policies that foreseeably produce the injustice and refuse to change the law to bring the injustice to an end have contribution-grounded remedial duties.

Having outlined the strong contribution principle, let me now highlight the problem with focusing exclusively on it to ground remedial duties. The problem is that there may be circumstances in which those who are identified as contributors by the

[7] These are similar to the conditions David Miller sets out for an agent to count as morally responsible (Miller 2007:86-90).

strong contribution principle are no longer around. To illustrate, return to the unjust society described just now. Imagine that all or almost of members of the governing regime die in a freak plane crash or are killed in a revolution. In this kind of case, those who count as contributors according to the strong contribution principle and so who are picked out as bearing remedial duties have left the scene entirely. What this shows is that, unless we are willing to say that remedial duties evaporate when contributors are not around, the strong contribution principle is insufficient for a full account of our remedial duties. The obvious way to address the problem is to appeal to additional principles to ground remedial duties, and this would, of course, be to take a multi-principle approach.

1.4. The associative principle

Let me now turn to the fourth principle I will examine: the *associative principle*. This principle holds that members of inherently valuable, non-voluntary social groups can owe each other special duties which they do not owe to non-group members. To be a member of such a group is to identify with it and recognise its inherent value, and the duties generated by these groups are constitutive of membership. Moreover, the specific content of the duties generated by each association will be determined by the particular values that characterise the group in question. It is widely accepted that the associative principle can generate obligations amongst the members of intrinsically valuable groups such as families, friends, and neighbours. I will assume that the associative principle can generate special duties amongst members of political communities. The plausibility of applying the associative principle to political communities derives from the observation that polities have considerable value and are typically taken to by their members (Horton 2007).

In invoking the associative principle to ground remedial duties, I am following Avia Pasternak (2010).[8] I should stress that while I am interested in how the associative

[8] Another theorist who grounds a duty to engage in political action in the associative principle is Michael Walzer. He observes that sub-state associations can sometimes impose obligations on their members that clash with the requirements of state citizenship. When this happens, Walzer argues that members of these sub-state associations are required to disobey the state with the aim of encouraging it to exempt their association from the requirement of citizenship in question (Walzer 1970:3-23). What is not clear is whether Walzer regards a situation in which the state and sub-state association are making competing and

principle might ground the duty to engage in social justice activism generally, Pasternak's specific focus is on the question of how the costs of remedying injustices perpetrated by democratic states should be distributed amongst the citizens of such states. She invokes the associative principle to defend an equal – as opposed to proportional in line with involvement – distribution of the costs of remedial efforts, arguing that such a distribution can be morally required in democracies characterised by an egalitarian, solidaristic ethos. Nevertheless, despite this difference in focus, it is clear that Pasternak thinks that the duty to contribute towards remedial efforts – however distributed – is one of the duties that can be generated by the associative principle when it is applied to valuable political communities.[9]

However, one might ask whether the associative principle can be invoked in a community characterised by social injustice at all. Defenders of the associative principle typically argue it cannot generate special duties amongst unjust groups. Pasternak, for instance, is explicit that her argument does not apply to groups that are "inherently bad or unjust" (Pasternak 2010:198). David Miller, another proponent of the associative principle, argues that membership in groups that "inherently involve injustice" cannot ground associative obligations. This is because "[t]he pervasive injustice [these groups] generate deprives them of such intrinsic value as they might otherwise have had" (Miller 2005:66).

Such caveats successfully rule out the generation of associative obligations in the most extreme cases – in Nazi Germany or apartheid South Africa, for instance – but not all social injustice is inherent to the political communities that practise them. Consider, for example, the fact that the Act of Parliament legalising same-sex marriage came into force in England and Wales on 13th March 2014. I believe that the fact that gay people were denied the right to marry before 13th March 2014 constituted a social injustice; gay people lacked equal and secure access to free association. It would seem a stretch,

---

incompatible demands on a group of individuals as a social injustice. It is therefore not clear that Walzer is defending a duty to remedy social injustice – that is, to engage in *social justice* activism.

[9] It is not clear whether Pasternak intends her account to be applied to cases of domestic injustice. She uses the example of the unjust occupation by Israel of the Palestinian Territories to illustrate her account, but, given the *de facto* authority Israel exercises over the Palestinians, it seems to me ambiguous whether the injustice here is domestic or international. In any case, nothing in Pasternak's account precludes applying it to clearer-cut cases of domestic social injustice, whatever her original intention.

however, to say that this injustice was inherent to the polity. After all, the legal reform that remedied the situation did not change the fundamental nature of the political community in the way that giving blacks the right to vote, say, transformed apartheid South Africa. Of course, it may be the case that, though not inherently unjust, the instances of injustice in any particular society far outweigh any value that it has. Associative obligations are likely to be undermined in such a scenario, though whether such a situation obtains will be a contested issue. The point I want to make here is that the mere fact that a political community contains some (non-inherent) injustice should not preclude the possibility that members of that community owe each other associative obligations.[10] Indeed, I think that one way to discharge associative obligations in conditions of (non-inherent) social injustice may be to contribute to remedying the injustice in question.

Let's assume that the associative principle can generate remedial duties in a society characterised by (non-inherent) injustice. To see the problem with focusing exclusively on the associative principle to ground remedial duties, recall that a key reason why the associative principle is said to generate special duties is that such duties are entailed by membership of a social group, and membership is itself constituted by identifying with the group and taking it to have inherent value. But this raises the obvious question: what about those who do not take the community to have such value? The answer, according to Pasternak, is that proponents of the associative principle "must concede that there may be some within the political community to whom the argument does not apply, namely, those who genuinely reject the claim that their political membership has some value" (Pasternak 2010:202, also see Renzo 2012).

The apparent implication of this for the duty to engage in social justice activism is deeply counterintuitive: those members of societies characterised by (non-inherent) injustice who do not see intrinsic value in their political community have no duties to remedy the social injustice in their midst, at least on associative grounds. This result is particularly troubling because those who see no inherent value in membership of society may be especially likely to happily contribute to or benefit from social injustice. One way

[10] As John Horton puts it: "although even the best polities will not be entirely just… and most will no doubt have a number of serious failings, this is not itself enough to show that people in such polities are therefore without any [associative] political obligations, so long as the polity has some value" (Horton 2006:439).

to avoid this conclusion would be to say that those who see no inherent value in the political community do have duties to remedy social injustice, but these duties are grounded in principles other than the associative principle. This seems to me to be a plausible move, but, clearly, it can only be taken as part of a multi-principle approach.

1.5. The principle of fairness

The *principle of fairness* holds that those who accept the benefits of a cooperative practice governed by rules are required to do their part according to the rules of the practice, so long as the practice itself is just (Rawls 1999a:96-8, 301-2). It seems plausible to suppose that the weight of each person's duty – what counts as 'doing their part' – will vary in line with the extent to which they benefit. The principle of fairness is widely accepted and is invoked to ground a variety of duties, ranging from the duty to keep promises to the duty to obey the law. Violating the principle of fairness – that is, accepting the benefits of an institution while failing to do one's part – is known as 'free riding'.

Candice Delmas has recently argued that the principle of fairness can ground a duty to remedy social injustice (Delmas 2014, 2018). Delmas establishes the case for this view in three stages. First, she argues that, because benefitting from unjust institutions involves the same kinds of wrongs that make free riding reprehensible, the principle of fairness prohibits benefitting from unjust institutions.[11] Second, Delmas argues that the sole way to cease benefitting from unjust institutions is to radically reform them. The final step in Delmas's argument is the claim that resistance – which in Delmas's view is a very broad category including any act that expresses opposition to the institution's rules – is typically necessary in order to bring about radical reform.[12] Delmas does not make

[11] In *A Duty to Resist*, Delmas expresses some ambivalence about whether the principle of fairness generates remedial duties for all beneficiaries of unjust institutions or only some (2018:117-123). Ultimately, she suggests that there will be instances when all those who benefit have fairness-grounded remedial duties, and she takes the Jim Crow system that existed in the southern USA to be an example of such as occasion. It seems to me that any situation in which an institution is generating benefits for some participants at the expense of others will be sufficiently like Jim Crow to suggest that, whenever the principle of fairness generates remedial duties for beneficiaries, it generates duties for all beneficiaries.

[12] Daniel Butt also argues that beneficiaries of injustice have duties to remedy the unjust state of affairs they face, but his claim that this conclusion is entailed by the very condemnation of a state of affairs as unjust is mysterious (Butt 2007:143-4). One way to understand Delmas's argument is as a demystification of this claim.

this point herself, but, it seems plausible to think that the weight of one's duty to remedy unjust institutions will vary in line with the extent to which one has benefitted from them.

I find Delmas's argument highly plausible. The problem with focusing exclusively on the principle of fairness to ground the duty to engage in social justice activism, however, is that it has nothing to say about the duties of non-beneficiaries. Delmas argues that both beneficiaries and victims (which do not exhaust non-beneficiaries) have fairness-based duties to join and support resistance movements which have already been initiated by victims and which are likely to be successful (Delmas 2018:129-133, also see Delmas 2014:484-5). But the principle of fairness is silent on the question of whether victims or other non-beneficiaries have duties to *initiate* (as opposed to join already existing) remedial efforts and, if so, what might *ground* such duties.

One might reply that victims of social injustice have been burdened quite enough and, therefore, have no duties to take on additional demands by engaging in social justice activism. This may be true for those victims who are below or at risk of falling below the minimum well-being threshold mentioned earlier. But not all victims will be in such a position. Some victims of injustice are more privileged than others and to assume that all victims have no duties in conditions of injustice is to treat them as passive patients in the project of remedying the social injustice they face, merely as recipients of the good deeds done to and for them by others. In other words, to suggest that all victims lack remedial duties simply because they are victims is, it seems to me, to deny or at least significantly downplay an important aspect of victims' moral agency.

My claim is not that Delmas is guilty of conceiving of victims in this way. In fact, as I will elaborate on in a moment, in her recent book, *A Duty to Resist*, Delmas highlights three other principles, in addition to the principle of fairness, which also ground remedial duties and which would pick out non-beneficiaries as duty-bearers. These principles are the natural duty of justice, the Samaritan principle, and the principle of political association. Hence, precisely because it does not focus exclusively on the principle of fairness, Delmas' view is not vulnerable to the charge that it is silent on question of whether non-beneficiaries have duties. In other word, hers is a multi-principle approach.

1.6. An intersectional approach

So far, I have examined five principles that seem to me to be plausible candidates for grounds of the duty to engage in social justice activism, identified what I take to be the

most plausible versions of each principle, and identified the problems associated with focusing on any of them exclusively when it comes to grounding remedial duties. To reiterate, my claim is not that many theorists defend a monistic approach to grounding remedial duties. Instead, my aim has been to make explicit a view that is often left unsaid or implicit: that we must take a pluralistic approach to grounding remedial duties. However, though this view is often not made explicit, I am not the first to take such an approach. It is, therefore, worth looking at other multi-principle or pluralistic approaches to grounding remedial duties to see how mine is distinctive.

1.6.1. Sequential, disjunctive, and intersectional approaches to grounding remedial duties

One example of a multi-principle approach to grounding the duty to engage in social justice activism is provided by Simon Caney, when he seeks to answer the question of who should remedy climate injustice. Caney notes that the principle often invoked to answer this question, the polluter pays principle, is limited by the fact that some harmful climate change stems from the contributions of people who no longer exist, from non-human activity, or from the contributions of those who are very poor and so should not bear further burdens. He labels such climate change "the Remainder" (Caney 2010:213). Caney argues that we should identify who should bear the burdens of dealing with climate change by, first, applying a qualified version of the polluter pays principle, and, then, invoking a qualified version of the ability to pay principle to deal with the Remainder. Caney's approach is pluralistic – he describes himself as defending a "hybrid" view (Caney 2010:218) – because it appeals to more than one principle. But it is also *sequential* in the sense that it advises that we start with one principle – the polluter pays principle – and then move to the other principle – the ability to pay principle – once the first principle has done all of the work it can.

David Miller provides another example of a pluralistic approach to grounding remedial duties. Miller is interested in finding the principle according to which agents can be assigned the duty to remedy situations in which people's basic rights are under threat. After surveying a range of candidate principles – causal responsibility, moral responsibility, capacity, and community – Miller concludes that no one principle is adequate on its own and so we must settle for a pluralistic approach. Unlike Caney, however, Miller does not think there is a particular sequence in which these principles should be invoked. As Miller puts it, "there is no algorithm that can tell us to apply

principle 1 first and then move on to principle 2, and so forth" (Miller 2001:471). Miller justifies this by focusing on the principle which he takes to be the most plausible candidate to be first in the sequence: moral responsibility. The problem, in Miller's view is that different agents will have varying amounts of moral responsibility, and the degree of moral responsibility will not necessarily correlate with other important considerations, such as capacity. "In these circumstances," Miller asks rhetorically, "is it right to let our judgements of remedial responsibility be determined entirely by our prior beliefs about how far different agents are morally responsible for creating the situation that needs a remedy?" (Miller 2001:467, also see Miller 2007:105-6). Instead, Miller argues that whenever there is a situation that needs remedying, we should assess the extent to which different principles connect different agents to the situation at hand and assign remedial duties according to whichever connection is stronger. We might say, then, that his pluralistic approach is *disjunctive* in the sense that it instructs us to apply *either* one principle *or* another.

The final pluralistic approach to grounding remedial duties I propose we look at is inspired by the work of both Candice Delmas and Iris Marion Young. As I have already mentioned, Delmas identifies four principles which she argues ground duties to remedy social injustice: the natural duty of justice, the principle of fairness, the Samaritan principle, and the principle of political association. There are three points worth highlighting about these principles (see Delmas 2018:224). First, different principles can generate the same obligations for the same people. There will be people who are picked out as bearing remedial duties by both the natural duty of justice and the principle of fairness, for example. Second, principles sometimes vary in terms of their coverage. We saw an example of this at the end of the previous subsection. Putting duties to join or support already existing remedial efforts to one side, the principle of fairness only generates remedial duties for beneficiaries of unjust institutions, but the natural duty of justice generates remedial duties for all, including victims and other non-beneficiaries. Third, principles can interact with one another. For instance, when discussing the Samaritan principle, Delmas invokes the principle of fairness to argue that all citizens should join remedial efforts grounded in the Samaritan principle (2018:162).

As I have also already noted, Young identifies what she takes to be a distinctive class of injustices which cannot be attributed to the wrongful actions of a limited set of specific agents, but is instead the result of large-scale social processes produced by the

actions of very many individuals. Participating in, and therefore producing and reproducing, these processes is not avoidable and so is not something for which agents can be blamed. Nevertheless, Young argues, all individuals who contribute to social injustice via these processes – including the victims of injustice – have a shared duty to transform them so that they no longer produce unjust results. The duties agents have are not identical, however. Instead the "kinds and degrees" (Young 2011:144) of agents' duties vary according to four additional considerations: the capacity agents have to influence the social processes that produce injustice; the extent to which they have benefited from the social processes that produce injustice; the level of interest agents have in the injustice being remedied; and the extent to which they are able to collaborate with others to remedy the injustice (Young 2011:144-147).

Both Delmas's and Young's approaches to social justice activism are pluralistic because they involve more than one principle. However, they differ from the other two pluralistic approaches I have mentioned[13] because, while Caney and Miller want whichever principle is applied to operate *independently*, Delmas and Young embrace the possibility that the principles they invoke may operate *concurrently and interactively*. Delmas and Young do not discuss their pluralistic approaches in much detail, but I find them both to be very suggestive. The pluralistic approach I defend is, in a sense, an attempt to bring to the fore and further develop some insights that are already present in Delmas's and Young's work.

I call my pluralistic account of the grounds of the duty to engage in social justice activism an *intersectional* approach because the principles appealed to do not operate independently of each other, but in combination with one another. In the paper in which she coins the term "intersectionality", Kimberlé Crenshaw argues that we cannot understand the experience of those who are multiply disadvantaged if we focus on each axis of discrimination in isolation; instead, we should adopt a framework that allows us to see how many different factors intersect (Crenshaw 1989). Similarly, I want to suggest that it is only by appreciating how different principles combine and interact that we can adequately explain the duties citizens have to engage in social justice activism. This is

[13] There is a sense in which Young's approach is, like Caney's, sequential. For Young, what triggers remedial duties for social injustice is contributing to the social structural processes that produce it. It is only once such duties have been triggered that we then move to the four additional principles to further specify the duties of particular agents.

certainly a repurposing of the framework of intersectionality: while Crenshaw uses an intersectional lens to analyse the impact of injustice on its victims, I deploy it to make sense of the remedial duties of all citizens. Nevertheless, I believe that such an extension is warranted and indeed called for. In one way or another and to varying degrees, most of us are victims of, participants in, and beneficiaries of some form of social injustice, after all.

According to the intersectional approach to social justice activism I favour, we should appeal to all of the five principles examined in sections 1.1-1.5 when grounding remedial duties. As well as avoiding the counterintuitive implications of invoking only one principle, such an approach also has an additional advantage. To see it, first notice that many of the principles for grounding the duty to engage in social justice activism we have just examined in the preceding sections will pick out different people as duty-bearers. The self-regarding principle will pick out a different group of citizens to the principle of fairness, for instance. Furthermore, some people will be picked out as duty-bearers by more than one principle. For example, those who count as contributors according to the strong contribution principle will also be picked out by the general, other-regarding principle. Let us characterise an agent who is identified as a duty-bearer by more than one principle as experiencing *overlap*.[14] It seems plausible to suppose that, holding all other things constant, the more overlap an agent experiences the weightier their duty to engage in social justice activism is; that is, the more costs that can be associated with the duty before it is extinguished. The one caveat to this would be that no duty can be so demanding that it requires people to fall below the minimum well-being threshold. It seems to me to be a considerable advantage of an intersectional approach that it is able to identify those agents who bear more or less weighty duties. In this way,

[14] This term is borrowed from George Klosko, who puts it to similar use in his account of the duty to obey the law (Klosko 2005:101). Similarly, Rawls argues that all citizens of a nearly just society have a duty (grounded in the natural duty of justice) to comply with and support the institutions of their state, but those who assume political offices have an additional (fairness-grounded) obligation to do so and so are "bound even more tightly" (Rawls 1999a:303). I take this to mean that those picked out as duty-bearers by both the natural duty of justice and the principle of fairness have weightier duties than those picked out by the natural duty of justice only. Delmas notes that the various principles she invokes to ground remedial duties can overlap, but does not draw the same implications from this that I do (Delmas 2018:224).

the intersectional approach enables us to make progress on question 3 identified in the introduction of this chapter, the weight question.

1.6.2. The intersectional approach: dealing with complexity

The intersectional approach is, then, clearly very complex. Even if you assume only one instance of social injustice obtains in society, the intersectional approach involves many different principles being invoked to ground the duties of members of society to engage in social justice activism. Different principles will apply to different sets of people, sometimes creating overlaps. Duties will vary in weight because some agents will experience more overlap than others. Things get more complicated once you notice that citizens will contribute to and benefit from injustice to varying degrees, and many instances of social injustice of varying severity will typically occur at any one time.

The picture is complicated even further when we consider a principle I have not looked at so far but may seem very relevant to the discussion of remedial duties: the *capacity principle*. According to this principle, it is those who are able to remedy social injustice who have duties to engage in social justice activism. I have not examined the capacity principle alongside the others because it does not seem to me to be what one might call a *primary grounding principle*. By this, I mean that the capacity principle cannot explain why the duty to remedy social injustice exists in the first place. Instead, it assumes that a bad situation should be remedied and asks: who should bear this duty? In this sense, the capacity principle presupposes that a primary grounding principle – the general, other-regarding principle, for instance – exists and has been triggered. Hence, we might say that the capacity principle is in fact a *secondary grounding principle*.

Moreover, as David Miller has noted:

> the capacity principle seems to blend together two different factors which may not always point in the same direction. One that has to do with the effectiveness of different agents in remedying the situation; the other has to do with the cost they must bear in the course of doing so. The strongest swimmer may also be so fearful (so that although he is an effective rescuer, the rescue causes him considerable distress) – or perhaps he simply dislikes the kind of attention that goes along with successful rescue. If A is slightly stronger than B, but A's costs are also much higher, is it obviously the right solution to hold A responsible for rescuing P?... [T]o apply the capacity principle, it seems, we have to begin by weighing effectiveness against cost to determine whose capacity is greatest in the morally relevant sense (Miller 2001:461).

The discussion so far has shown that there is no way to 'determine whose capacity is greatest in the morally relevant sense'. This is because it is not possible to speak generically of the costs people should bear when engaging in social justice activism. The costs that people should bear are determined by the weight of their duty and, because of overlap (amongst other things), the weight of agents' duties can vary.

There is no doubt that this picture is messy and somewhat bewildering. I do not think this alone is a reason to reject it, however; quite the opposite, in fact. Social life is complicated. Morality contains many principles for action. Political communities typically contain a number of social injustices and different people will stand in a variety of relations to any particular instance of injustice and to others who encounter it. Any theory of social justice activism that is too neat and tidy is unlikely to adequately reflect the circumstances that characterise the real world. The fact that the approach I present is complex is, therefore, entirely appropriate and to be expected.

Nonetheless, while it is true that morality is complex, I take it that it would count against the intersectional approach I propose if it was not able to help citizens orient themselves in the midst of all this bewildering complexity, if it just insisted that morality is complex and then left it to citizens to figure the rest out for themselves. Though I think there is a limit to how much any theory can prescribe, I also think that the intersectional approach may be slightly more helpful than that. In particular, I think that the progress that the intersectional approach enables us to make on question 3 identified in the introduction of this chapter – the weight question – may also allow us to make some modest progress on question 2 identified in the introduction of this chapter – the content question.

To see how this is so, we should first make the fairly straightforward observation that, when a citizen is deliberating about how to discharge her remedial duty, she is faced with a vast array of options. There are numerous ways in which these options might be distinguished from each other, but one way is according to the level of cost or burden associated with engaging in different forms of activism. Costs can take an extremely wide range of forms, including volunteering time, spending money, enduring social disapproval, or even being subject to physical harm. Citizens will vary in terms of their subjective evaluations of different experiences such that the same situation or experience will be regarded as differently costly by different people. Some people are extremely

sensitive to social disapproval, for instance, while others are able to withstand it easily. However, as I mentioned earlier, I assume an objective understanding of cost such that there are certain experiences that qualify as costs for all people, regardless of any given person's subjective attitude towards these experiences.

On this basis, I think it is possible to identify forms of activism that fall into the following very broad categories: low-cost, mid-cost, and high-cost. Examples of low-cost social justice activism include voting for the candidate or political party that is likely to do the most to remedy social injustice, or joining a peaceful demonstration in support of a policy likely to remedy or ameliorate social injustice. Examples of mid-cost social justice activism include donating a modest but not insignificant proportion of your income or wealth to the cause of remedying social injustice, or speaking out against injustice in a way that means you forego some social approval. Examples of high-cost social justice activism include volunteering the majority of your time and efforts to an organisation that seeks to dismantle social injustice, or, in the most extreme case, engaging in forms of activism that carry a nontrivial risk of severe physical harm.

The intersectional approach cannot tell citizens exactly what they must do to remedy social injustice; I don't believe any theory could do that. But it can help to narrow the range of options that a citizen should choose from by indicating which cost category their remedial duties are likely to fall into. *All other things being equal*, I want to suggest that an individual citizen can roughly approximate whether her remedial duties are low-, mid-, or high-cost by counting the number of principles that pick her out as a duty-bearer. A citizen picked out by one principle is likely to have low-cost remedial duties, and the only way to be in such a position is to be picked out by the general, other-regarding principle but none of the others. A citizen picked out as a duty-bearer by 2 or 3 principles is likely to have mid-cost remedial duties, and there are a wide variety of combinations of principles that might produce this kind of situation. A citizen picked out as a duty-bearer by 4 or more principles, however, may have high-cost remedial duties. Again, there are a wide variety of combinations of principles that might produce this kind of situation but, given that all citizens will be picked out as duty-bearers by the general, other-regarding principle, every combination will, in addition, include at least three of the remaining four principles: the self-regarding principle, the contribution principle, the associative principle, and the principle of fairness.

It is worth stressing the importance of the 'all other things being equal' condition attached to this suggestion. For the purposes of this discussion, it means that we are focusing on a single case of social injustice and we are assuming that the extent to which anyone contributes to or benefits from this injustice, or identifies with one's community, etc. is the same. Variation in the degree to which citizens contribute to or benefit from social injustice or identify with their community (etc.) leads to variation in the weight of citizens' duties, and accommodating the likelihood of multiple instances of social injustice of varying severity complicates the picture further still. It goes without saying that, in reality, all things are rarely equal, but the purpose of the condition is to help illustrate the implications of the intersectional approach most clearly.

1.7. Conclusion

In this chapter, I have argued that we should invoke multiple principles when grounding remedial duties, suggested that these principles can combine and interact, and highlighted how taking such an approach can help us to ascertain the weight and, to a lesser extent, the content of citizens' duties to engage in social justice activism. In order to make more progress on the content question, we must be clearer about the different forms that social injustice can take. I explore this issue in the next chapter.

2. Three sites of social justice activism

> Understanding the nature of collective behaviours and why people engage in them is crucial for the design of interventions aimed at social change.
>
> Bicchieri 2017:1

Introduction

My aim in this dissertation is to develop a theory of how to remedy social injustice – that is, a theory of social justice activism. In the previous chapter, I argued for what I called an intersectional approach to grounding the duty to engage in social justice activism. But, in addition to the issue of the grounds of remedial duties, a key question for a theory of social justice activism relates to the content of the duty to remedy social injustice. Put simply, the question is: how should social justice activism be practised?

Social injustice obtains because of social practices – that is, the patterns of behaviour that are widespread in society, including, of course, practices backed by law. This means that, in order to know how to discharge our duties to remedy social injustice, we must understand the nature of these social practices and the mechanisms that underpin them. My primary aim in this chapter is to highlight three kinds of social phenomena that can help us to explain social injustice. I do not claim that these are the only social phenomena that can generate social injustice or that they typically operate independently of each other. Nevertheless, I focus on these mechanisms because I think they are particularly important, and I discuss them separately so as to bring their distinctive features more clearly into view. I refer to these social phenomena as the 'three sites of social justice activism' because I will explore the ethics of activism that seeks to intervene in each of them in the chapters to come.

This chapter proceeds as follows. In sections 2.1-2.3, I introduce culture, informal rules, and formal rules as three categories of social phenomena. These categories stand in a particular relation to each other: culture is a very capacious category that includes social rules, and social rules can be divided into informal and formal varieties. I focus on three specific instantiations of culture, informal rules, and formal rules as mechanisms that can generate social injustice: respectively, stereotypes, social norms, and law. It is these three mechanisms that I am referring to when I speak of sites of social justice

activism, and I provide examples to illustrate how social injustice can be generated by each one. In section 2.4, I highlight some of the ethical complexities and controversies associated with intervening in each of the specific social phenomena I introduce, thereby providing a preview of the discussions which take place in the chapters that follow. Section 2.5 concludes.

It is worth noting that, while sections 2.1-2.3 broadly follow existing accounts of the three specific social phenomena just mentioned, in sections 2.2 and 2.3, I also argue for amendments to and an integration of existing accounts of social norms and law. Thus, as well as laying the foundations for the chapters to come, an additional aim of this chapter is to contribute to debates in social ontology.

2.1. Culture: stereotypes

The first category I will introduce is culture. I follow Dan Sperber in defining culture in terms of *representations* – that is, objects that represent something for human beings. Sperber distinguishes between mental and public representations. *Mental representations* include memories, beliefs, and intentions, and reside in the minds of individuals. *Public representations* are features of the environment individuals occupy – e.g., symbols, texts, utterances – and are often the means via which individuals communicate their mental representations to others, for these others to then transform into mental representations for themselves. Some mental representations are fleeting but some are more durable. And some of these more durable representations are communicated repeatedly, so versions of them end up in the minds of very many individuals. *Culture*, according to Sperber, refers to widely distributed, long-lasting mental representations, and also to the public representations via which mental representations are propagated (Sperber 1996).

Culture, therefore, covers an extremely wide range of phenomena, from cooking recipes, to scientific theories, from religious beliefs, to artefacts, practices, etc. Indeed, all the social phenomena I introduce in this chapter – including those I will come to discuss in sections 2.2 and 2.3 – are included amongst culture, according to the definition just given. The instantiation of culture that I would like to highlight in this section is a stereotype. I focus on stereotypes because they can generate social injustice in a distinctive way, which I think demonstrates the pervasive influence of culture on behaviour.

A *stereotype* is a widely distributed representation of a social group that helps those who possess it to make sense of their social environment by simplifying the characteristics associated with the group being represented (McGarty et al. 2002:1-7). Stereotypes are often negative – for instance, in many western societies, black people are associated with unpleasant characteristics such as 'unintelligent' and 'violent' – and so they are sometimes explicitly rejected by the members of the societies in which they are widespread. Because of this, various tools have been devised to measure the prevalence of stereotypes without having to ask participants to report on the extent to which they identify with or endorse the stereotype in question. Perhaps the most well-known and widely used of these tools is the "Implicit Association Test" (IAT hereafter, Greenwald et al. 1998).

The IAT involves participants sorting stimuli (e.g., black- or white-sounding names, and pleasant and unpleasant words) into categories (e.g., black or white, pleasant or unpleasant). Participants are asked to sort the stimuli into their correct categories as fast as possible, and categories are paired so that they are either consistent (white and pleasant, black and unpleasant) or inconsistent (white and unpleasant, black and pleasant) with common stereotypes. The IAT measures the speed and accuracy with which participants sort stimuli into categories when those categories are paired in stereotype-consistent versus stereotype-inconsistent ways. Being faster and more accurate at sorting stimuli when the categories are paired in stereotype-consistent versus inconsistent ways is interpreted as evidence that one has imbibed or absorbed[15] the stereotype more strongly. And the extent to which one has absorbed a given stereotype, as measured by tools such as the IAT, is often referred to as the extent of one's *implicit bias* in favour of the stereotype. Decades of research using tools such as the IAT have shown that measures of implicit bias in favour of a given stereotype are not strongly correlated with the responses participants give when asked about the extent to which they identify with or endorse that stereotype (see Brownstein 2019:section1.2).

It is fairly clear how widely endorsed negative stereotypes might underpin the kinds of discriminatory patterns of behaviour that social injustice often consists of. To

[15] In general, while the process of imbibing or absorbing something is often unintentional or unconscious, this need not necessarily be the case. One might visit a foreign country with the explicit intention of imbibing the language or culture of that country, and one might monitor one's progress and so be quite conscious of the process as it is happening.

illustrate how imbibed – though not necessarily endorsed – stereotypes might generate social injustice, consider the following two examples. First, various studies conducted in different countries have found that job applications from candidates with white-soundings names are significantly more likely to receive call-backs for interviews compared to equivalent applications from candidates with African American- or ethnic minority-sounding names (Bertrand and Mullainathan 2004, Wood et al. 2009, Rooth 2010). One study in particular found that applicants with Arab-Muslim-sounding male names are less likely to receive a call-back from Swedish recruiters than applicants with Swedish-sounding male names. Moreover, this same study also found that most Swedish recruiters harbour implicit biases in favour of negative stereotypes of Arab-Muslim males, and many Swedish recruiters explicitly report that they would prefer to hire a native Swedish male over an Arab-Muslim male. Crucially, however, while harbouring negative implicit biases about Arab-Muslim males is strongly correlated with a lower probability of calling back candidates with Arab-Muslim-sounding names, there is no such correlation between explicit prejudice and call-backs (Rooth 2010).

For a second example, consider the fact that women who are regarded as successful in jobs stereotyped as male tend to be regarded as less likeable, and, as a result, tend to be evaluated less favourably in the workplace or as prospective employees (Fiske et al. 1991, Rudman and Glick 2001, Eagly and Karau 2002, Heilman et al. 2004). For instance, one study asked participants to assess equivalent applications for a managerial position from a male and a female candidate, where both presented as stereotypically masculine (e.g., competitive, independent, hierarchical). The researchers found that, while participants rated both applicants as equally competent, they also perceived the female candidate to be less socially skilled and, therefore, to be less hireable. Importantly, harbouring an implicit bias in favour of stereotypical representations of men (competitive, independent, hierarchical) and women (communal, cooperative, kind) was correlated with rating the female candidate as less socially skilled, while endorsement of these stereotypes was not (Rudman and Glick 2001).

These examples show that discriminatory behaviour is sometimes better explained by the extent to which a stereotype has been absorbed, rather than the extent to which that stereotype has been endorsed. This is a rather subtle and mysterious way for social injustice to be generated and it is likely that, in order to remedy social injustice caused in this way, we need to better understand how stereotypes work and the

distinctive challenges they throw up for social justice activism. I will highlight some of these challenges in section 2.4.

2.2. Informal social rules: social norms

The next two categories that I will introduce are varieties of social rules. A *rule* is a proposition that expresses a general standard of behaviour; it is general in the sense that, unlike a one-off instruction or order, it requires the relevant behaviour whenever a given occasion or type of occasion arises. A *social rule*, as I understand it, is social in two ways. First, it is a rule that sets a general standard of behaviour for a group of people. Thus, it differs from what we might call a *private rule*, which applies to one person only. The second way that a social rule is social is that it manifests itself as a widespread and recurring mental representation within the group of people to whom it applies. So, for instance, the proposition 'thou shall not take the lord's name in vain' is a rule because it expresses a general standard of behaviour, and it is a social rule because it applies to more than one person and is regularly recurring in the minds of many of those to whom it is said to apply. To the extent that a social rule is a widespread and enduring mental representation, it is also a feature of culture.[16]

Following H. L. A. Hart, it is worth distinguishing between two kinds of social rules: primary and secondary. *Primary (social) rules*[17] directly concern the conduct of individuals, requiring them to perform or refrain from performing particular actions, whether they want to or not. *Secondary rules*, on the other hand, "are all concerned with the primary rules themselves. They specify the ways in which the primary rules may be ascertained, introduced, eliminated, varied, and the fact of their violation conclusively determined" (Hart 2012:94). Sometimes rules exist as part of a network of rules that includes secondary rules about how the rules of that network are identified, created,

[16] A rule can also be a feature of culture if it is manifested as multiple and/or widely distributed public representations such as the text of books, even if it is rarely or never manifested as mental representations in the minds of individuals. We might call this a *dormant element of culture*, to be contrasted with a *living element of culture*.

[17] My focus in this chapter is social rules in particular not rules in general. However, because I will identify various categories of social rule, I will often omit the 'social' for ease of exposition. For the remainder of this chapter, whenever I speak simply of rules I will be referring to social rules specifically. Hart, for his part, uses the terms 'rule' and 'social rule' interchangeably, though his definition is different from mine.

modified, applied, interpreted, and enforced. Let us call rules that are part of such a network *formal rules* and rules that are not part of such a network *informal rules* (see Brennan et al. 2012:40-56).[18] I discuss an example of an informal rule in the rest of this section and an example of a formal rule in the next.

The example of an informal rule that I focus on is a social norm and I subscribe to Cristina Bicchieri's definition of social norms. Bicchieri's account is complex and the following ideas are crucial to understanding it: conditional preferences, empirical expectations, normative expectations, reference networks, and personal normative beliefs. *Preferences*, according to Bicchieri, "are dispositions to act in a particular way in a specific situation" (Bicchieri 2017:6, also see Bicchieri 2006:6), and preferences are *conditional* when this disposition to act in a given way depends on something else being the case. So, to say that I have a conditional preference for queuing is to say that I am disposed to queue not whenever I must wait to be served (i.e., unconditionally) but only when a certain state of affairs obtains – when other people are queuing, for instance.

Expectations are beliefs about what is going to happen in the future, and Bicchieri distinguishes between two kinds: empirical and normative. *Empirical expectations* are beliefs about what other people do. So, the belief that most other people will queue is an empirical expectation. *Normative expectations* are beliefs about what other people believe should be done; they are, therefore, second-order beliefs – beliefs about beliefs. So, the belief that others believe that people should queue is a normative expectation. Bicchieri refers to empirical and normative expectations together as "social expectations" (Bicchieri 2017:11-14). A *reference network* is the group of others whose behaviour and beliefs social expectations refer to (Bicchieri 2017:14). So, when it comes to queuing, one's reference network might be one's household, wider family, local community, or nation.

Finally, *personal normative beliefs* are the first-order beliefs concerning what should be done that normative expectations are about. These beliefs are not merely prudential, according to Bicchieri. Instead, they are more strongly normative beliefs about rules people ought to comply with, whereby noncompliance with the rule in question elicits disapproval and the likelihood of sanctions (Bicchieri 2017:35, also see Bicchieri 2006:14).

[18] Brennan et al. use the label 'non-formal' instead of 'informal'.

Using these key ideas as building blocks, Bicchieri defines a social norm as

> a rule of behaviour such that individuals prefer to conform to it on condition that they believe that (a) most people in their reference network conform to it (empirical expectation), and (b) that most people in their reference network believe they ought to conform to it (normative expectation) (Bicchieri 2017:35).[19]

Bicchieri describes her definition of a social norm as a "rational reconstruction" (Bicchieri 2006:3, 10). By this, she means that her account should not be taken as a faithful description of the beliefs and desires people consciously have, but as a way of uncovering and specifying the basic elements and architecture of social norms which typically lie outside of conscious awareness. Crucially, according to Bicchieri, a successful rational reconstruction is one that is both explanatory, in the sense that it can help to make sense of the empirical results of experimental studies, and testable, in the sense that it identifies the precise features of social norms which can then be manipulated to see whether behaviour does indeed change in predictable ways. Much of Bicchieri's work has been dedicated to showing that this is exactly what her definition of social norms provides (e.g., Bicchieri 2006:100-213, Bicchieri et al. 2018:section 6).

There are two features of Bicchieri's account that I think it is worth emphasising. The first is that a rule can be complied with as a social norm even though one or both of the beliefs upon which compliance depends is false. The belief set out in (a) in the passage from Bicchieri quoted above might be false in the case of social norms regulating private behaviour. For instance, someone might interpret the fact that there are few premarital pregnancies as evidence of the fact that most others are complying with a rule proscribing premarital sex, when this rule is in fact widely violated (Bicchieri 2006:13). If the belief set out in (b) is false, then it is possible that none of those who follow the rule have personal normative beliefs that correspond to it. People follow the rule in part because they mistakenly believe that others believe they should. Examples include gang members who are privately uncomfortable about their violent, antisocial behaviour, but do not

[19] Bicchieri presents her account of social norms in different ways in different places (see, e.g., Bicchieri 2006:11 and Bicchieri 2017:65-6, which are both different from each other and from the definition quoted in the text). While I take these different presentations to be consistent with each other, I have chosen to quote this presentation because it is recent and seems to me to be the clearest.

express these concerns in public and so appear to other gang members as fully committed, or prison guards who have liberal private attitudes towards disciplining inmates, but falsely assume that these attitudes are not widely shared (Bicchieri 2006:184). Such situations are characterised by a phenomenon known as *pluralistic ignorance*: "a cognitive state in which each member of a group believes her personal normative beliefs and preferences are different from those of similarly situated others, even if public behaviour is identical" (Bicchieri 2017:42 reference removed).

The second feature worth emphasising about Bicchieri's account of social norms is that it is an implication of her definition that a social norm can exist but not be complied with. This would be the case if individuals preferred to comply with a rule conditional on having the beliefs set out in (a) and (b) above, but these conditions for compliance are not in fact met (Bicchieri 2017:66, also see Bicchieri 2006:11).

I suggest that we systematise these features of Bicchieri's account by distinguishing between the following three kinds of social norm. The first kind exists when a rule is followed and the beliefs upon which compliance depends are correct. Let us call this a *robust social norm*, to reflect the fact that compliance with the rule is durable, that is, likely to endure or persist. The second kind of social norm exists when a rule is followed and at least one of the beliefs upon which compliance depends is false. Let us call this a *fragile social norm* to reflect the fact that compliance with the rule is vulnerable to collapse if people become aware that the beliefs they currently hold are incorrect.[20] The third kind of social norm exists when a rule is not followed but it would be if the relevant beliefs were present. Let us call this a *latent social norm* to reflect the fact that compliance with the rule would emerge if people's beliefs were to change.[21] It is worth noting that the same rule can exist as a different kind of social norm for different individuals in the same society. This would be the case if different individuals have different reference networks and the behaviour and beliefs that compliance with social norms depends on are not uniform across these networks. Figure 1 below presents this analysis of the three kinds of social norms and the questions one should ask in order to identify them.

[20] Bicchieri herself describes social norms complied with on the basis of mistaken beliefs as "fragile" and vulnerable to collapse if the veracity of these beliefs was revealed (Bicchieri 2006:195), but she does not incorporate this analysis into her definition of social norms.

[21] This choice of labelling is inspired by Spiekermann (2015:177).

Figure 1: Identifying three kinds of social norm

Do individuals have preferences to comply conditional on social expectations?	Yes	Yes	Yes
Do individuals have the relevant social expectations?	Yes	Yes	No
Are the social expectations true?	Yes	No	N/A
	Robust social norm	**Fragile social norm**	**Latent social norm**

Let me provide an example to illustrate how social norms can help us to explain social injustice, and to demonstrate the usefulness of my proposal to disaggregate and re-label Bicchieri's account of social norms. Consider the election of Donald Trump as the President of the USA in 2016, and the rule that one should not denigrate or abuse people who are members of disadvantaged groups in society. During his campaign for the Presidency, Trump continually demonstrated that he does not comply with or endorse this non-abuse rule, and, according to analysis of FBI data, Trump's election "appears to have contributed to a statistically significant uptick in hate crimes" in the USA (Edwards and Rushin 2018:13).[22] Moreover, the same analysis suggests that "at the county level, strong support for Trump during the presidential election was associated with a larger and more statistically significant uptick in the number of reported hate crimes" (Edwards and Rushin 2018:16).

It seems that a plausible explanation of this sequence of events is as follows. The rule that one should not denigrate or abuse people who are members of disadvantaged groups in society is a social norm for many Americans. That is, many Americans prefer to comply with this rule on the condition that they believe that (a) most people in their reference network comply with it and (b) most people in their reference network believe they ought to comply with it. However, the fact that Trump was elected after running such an offensive campaign revealed to many Americans that far fewer people endorse the

[22] In fact, according to Griffin Edwards and Stephen Rushin, "Trump's rise to the Presidency was associated with one of the largest upticks in hate crime in recorded American history – second only to the spike in hate crimes after the terrorist attacks on September 11, 2001" (Edwards and Rushin 2018:3).

non-abuse rule than they previously believed. Moreover, because reference networks are often local, the change in normative expectations was especially dramatic in areas where support for Trump was strongest.

The result is that there are people – especially in counties where support for Trump was strong – who previously complied with the non-abuse rule who no longer do so, because they no longer believe that most people in their reference network endorse the rule.[23] Effectively, the election of Donald Trump as President has highlighted that the non-abuse rule is not only a social norm for many Americans but was in fact a fragile social norm for some prior to 8th November 2016. For those who no longer comply with the non-abuse rule but would if they believed most of their (local) reference network did and believed they should, the non-abuse rule is now a latent social norm. On the other hand, the non-abuse rule remains a robust social norm for many of those who live in counties where support for Trump was weaker.

My proposal to amend Bicchieri's account of social norms is a very friendly one. It is simply a matter of re-labelling and making explicit categories and insights that were already implicit within it. Nevertheless, as the example of Trump's election and the rise in hate crime demonstrates, being clearer about the different forms social norms can take helps us to make sense of the uneven and dynamic nature of compliance with social norms. More generally, the fact that social norms can underpin social injustice means that they are a site of social justice activism – that is, a social phenomenon that those seeking to remedy social injustice will need to intervene in. I will highlight some of the ethical complexities and controversies associated with intervening in social norms in section 2.4.

2.3. Formal rules: law

Having introduced informal rules as category of social phenomena and social norms as a specific site of social justice activism, let me now turn to formal rules. Recall that a rule is formal when it is part of a network of rules that includes secondary rules about how the rules of the network are identified, created, modified, applied, interpreted, and enforced

[23] This is consistent with Edwards and Rushin's explanation of the data. They argue that "the election of Donald Trump may have validated his campaign rhetoric. It may have signalled to would-be hate crime perpetrators that the grievances President Trump raised in the campaign… were shared by a significant cross-section of the American population (Edwards and Rushin 2018:20).

(Brennan et al. 2012:40-55). The paradigmatic example of a formal rule according to this definition is law as analysed by H. L. A. Hart in *The Concept of Law*. There, Hart rejects what he takes to be the prevailing understanding of law as primary rules issued by a sovereign, on the ground that such a view cannot explain the most basic and familiar features of a domestic legal system (Hart 2012:26-80). According to Hart, the best way to understand law is by paying attention to *both* primary and secondary rules and the interplay between them. As he puts it: "[w]e accord this union of elements a central place because of their explanatory power in elucidating the concepts that constitute the framework of legal thought" (Hart 2012:81).[24]

Hart's view is that, for there to be law, it is not enough that the citizens of the society comply with primary rules issued by the sovereign. Instead, it must also be the case that a special class of individuals complies with a set of secondary rules that are connected to the primary rules of the system in specific ways. The special class of individuals Hart has in mind are the society's officials. Hart does not explicitly define what it means to be an official, but it seems to me that the most plausible way to do so would be to borrow from Max Weber and characterise them as those members of society who successfully lay claim to the monopoly of legitimate violence (Weber 1994[1919]:310-311).

There are three kinds of secondary rule in the set that officials comply with, and they each serve a distinct purpose. The purpose of the first rule is to remove uncertainty around how to identify those primary rules that are to count as laws, by specifying the features that constitute conclusive proof that a given primary rule is in fact a law. Hart calls this rule the "rule of recognition" (Hart 2012:94). The purpose of the second kind of rule is to facilitate change to the laws of the system by specifying the procedure via which laws are to be introduced and eliminated. Hart calls these rules "rules of change" (Hart 2012:95). The purpose of the third kind of rule is to rationalise and centralise the process by which violation of the law is identified and punishment for violation is enforced. Hart calls these rules "rules of adjudication" (Hart 2012:97). According to Hart, "law may most

[24] It is worth noting that both Hart and Bicchieri invoke the ability to explain relevant empirically observable phenomena as the standard against which to measure the success of their accounts of law and social norms respectively.

illuminatingly be characterized as a union of primary rules of obligation with such secondary rules" (Hart 2012:94).

For a legal system to exist, on Hart's view, officials must not simply comply with these secondary rules; they must also regard them from what Hart calls the "internal point of view". This means that the officials do not look upon these rules as if they were external observers seeking to record how other people behave and to identify – so as to avoid – the behaviour likely to trigger sanctions. Instead, officials *accept* the secondary rules of recognition, change, and adjudication as standards of behaviour that officials *should* follow. They do not merely regard failure to comply with these rules as something likely to trigger sanctions but as a *good reason* for those sanctions (Hart 2012: 55-7, 89-90). For Hart, officials viewing the secondary rules they comply with from the internal point of view "is logically a necessary condition of our ability to speak of a single legal system" (Hart 2012:116). Indeed, Hart suggests that these two elements – general compliance and being viewed from the internal point of view – are necessary for us to be able to say that any kind of rule, whether primary or secondary, exists within a group (Hart 2012:54-8, 85).

We are very familiar with social injustice generated by law and activism that seeks to intervene in law to remedy such injustice. Examples of injustice include laws which prohibit sexual relations between consenting adults of the same sex, laws and policies which severely limit the reproductive or economic freedoms and opportunities open to women, or laws which deny political rights to members of an unfavoured race. And corresponding examples of activism include the campaigns to abolish such laws. Hart's account of law – in addition to being highly illuminating on its own terms – brings some of the ethical complexities and controversies associated with intervening in law to remedy social injustice more clearly into view. I will outline these complexities and controversies in section 2.4.

2.3.1. Integrating Hart's account of law and Bicchieri's account of social norms

Though I find Hart's account plausible and appealing, I also want to suggest that it be amended. More specifically, I want to suggest that we extend Bicchieri's account of social norms to supplement Hart's theory by conceiving of the secondary rules that officials comply with as social norms as Bicchieri defines them. That is, I want to suggest that the rules of recognition, change, and adjudication are best understood as rules that officials

comply with because they believe that: (a) most officials comply with these rules and (b) most officials believe they (that is, officials) should comply with the rules.

Bicchieri's conception of social norms is in some ways consistent with how Hart characterises officials' relationships to secondary rules, but in other ways it is not. The disaggregation and re-labelling of Bicchieri's account which I suggested in the previous section actually helps us to see this. Recall that I suggested that we distinguish between robust, fragile, and latent social norms. A robust social norm is a rule a person follows because they *correctly* believe that most members of their reference network: (a) comply with the rule and (b) believe that members of their reference network should comply with the rule. So, a robust social norm is a rule that is actually complied with and which actually has personal normative beliefs that correspond to it. This is completely consistent with the existence conditions that Hart identifies for rules in general and so for secondary rules in particular: general compliance with the rules and viewing the rules from the internal point of view. Indeed, regarding a rule from what Hart calls the internal point of view is equivalent to having what Bicchieri calls personal normative beliefs that correspond to it.

The inconsistencies between Hart's account of law and Bicchieri's account of social norms arise because Bicchieri does not regard widespread compliance or corresponding personal normative beliefs to be necessary conditions for the existence of social norms, while Hart does take these features to be necessary conditions for the existence of secondary rules (and rules in general). As we saw in section 2.2, Bicchieri's definition of social norms allows for cases in which a social norm exists but is not complied with. Hart assumes that compliance is necessary for existence. Moreover, for Bicchieri, having personal normative beliefs that correspond to a social norm is not necessary for compliance with it. Compliance with social norms depends on normative expectations – beliefs about the presence of personal normative beliefs – and these expectations can be mistaken, as in the case of pluralistic ignorance. For Hart, on the other hand, officials must have personal normative beliefs that correspond to the secondary rules of recognition, change, and adjudication – that is, that they must regard these rules from the internal point view.

Despite these inconsistencies, I want to suggest that the officials of a legal system are best understood as complying with the secondary rules of recognition, change, and adjudication as social norms as Bicchieri defines them. This is no doubt an amendment

to Hart's view. Nevertheless, I will try to show that such an amendment is consistent with some of Hart's other comments and concerns. Moreover – and in any case – I will argue that it is warranted as it enhances the explanatory power of Hart's account.

To see how my proposed amendment can be deemed consistent with some of Hart's comments and concerns we must keep two things in mind. First, we must recall my suggested disaggregation and re-labelling of social norms as Bicchieri defines them, which generates the categories of robust, fragile, and latent social norms. Second, we must briefly examine Hart's discussion of what he calls the "pathology and embryology of legal systems" (Hart 2012:122).

According to Hart, in the "normal, unproblematic case where we can say confidently that a legal system exists… the rules recognized as valid at the official level are generally obeyed" (Hart 2012:117-8). But reality can deviate from this standard in a number of ways, and Hart discusses three. The first is when "the official sector [becomes] detached from the private sector, in the sense that there is no longer general obedience to the rules which are valid according to the criteria of validity in use in the courts" (Hart 2012:118). Examples of such a case include revolution, occupation, and anarchy.

The second is when "a new legal system emerges from the womb of an old one"; Hart cites the experience of the former British colonies as an example of this. In this kind of case, there is a process which begins with one legal system: the officials of the colony regard the rule of recognition of the United Kingdom Parliament as the ultimate criterion for identifying the laws of the colony. At the end of the process, however, there are two independent legal systems. This is because

> [t]he legal system in the former colony now has a 'local root' in that the rule of recognition specifying the ultimate criteria of legal validity no longer refers to enactments of another territory. The new rule rests simply on the fact that it is accepted and used as such a rule in the judicial and other official operations of a local system whose rules are generally obeyed (Hart 2012:120).

The third way for reality to deviate from the 'normal, unproblematic case' is for there to be division amongst the officials about the secondary rules of the system. As an example, Hart cites "constitutional troubles in South Africa in 1954". In this case, the legislature and the courts came into a conflict over a criterion of legal validity, the legislature created "a special appellate 'court' to hear appeals from judgments of the ordinary courts", with the special court reversing the judgments of the ordinary courts

and the ordinary courts declaring the special court invalid. The process was eventually stopped "because the Government found it unwise to pursue this means of getting its way" (Hart 2012:122).

There are two things that are particularly interesting about these cases in light of our current discussion. The first is that the second and third cases involve the deviations from the 'normal, unproblematic case' which refer to officials' attitudes and behaviour. In the case of a new legal system emerging from an old one, the officials in the colony transition from accepting one set of secondary rules – regarding them from the internal point of view – to accepting another. Presumably this is a departure from the 'normal, unproblematic case' because there will be a point when at least some officials no longer accept the outgoing secondary rules but still comply with them, or begin to comply with the incoming rules but don't yet accept them. Similarly, in the case of officials who are divided over a criterion of legal validity, there is a moment when officials of the same system are complying with competing and incompatible secondary rules. The second point that is interesting about the examples of deviations from the 'normal, unproblematic case' Hart cites is that he explicitly cautions against too swiftly drawing the conclusion that a legal system has ceased to exist. This is because the statement that a legal system exists is "broad", "general", and "elastic" enough to accommodate deviations from the 'normal, unproblematic case' (Hart 2012:118, 123).

Turning back to the three kinds of social norms, I have already noted that officials complying with the secondary rules of recognition, change, and adjudication as robust social norms perfectly satisfies one of the criteria Hart sets out for the 'normal, unproblematic case': officials regard the rules they comply with from the internal point of view. Let us call a legal system in which officials comply with the secondary rules as robust social norms a *robust legal system*.[25] Such a system is the 'normal, unproblematic case' and, other things being equal, is most likely to persist.

Moreover, it also seems to me that the categories of fragile and latent social norms are apt to make sense of some of the ways that reality can deviate from the 'normal, unproblematic case'. To illustrate this with respect to fragile social norms, imagine a deeply oppressive society and legal system in which all opposition and dissent is brutally

[25] I'm assuming here that the other feature of the 'normal, unproblematic case' is also in place: citizens generally obey the laws the officials deem valid.

crushed. Initially, the officials accept the secondary rules of recognition, change, and adjudication of the legal system and regard them from the internal point of view. But, over time, they begin to lose faith with the system and eventually they cease to accept its secondary rules. However, because of the oppressive nature of the system and the fear of the severe punishment that deviance may trigger, none of the officials voices their misgivings. Instead, they continue to go through all of the motions and perform all of the operations required to keep the system going. Indeed, they do so enthusiastically in an effort to disguise the fact that they have had a change of heart about the system.

Such a system is one characterised by pluralistic ignorance amongst the officials and in which the secondary rules are fragile social norms: the officials comply with the rules because they mistakenly believe that other officials believe they should comply. This kind of legal system is vulnerable to undergo a rapid and unexpected breakdown if officials' true personal normative beliefs are revealed to each other, and something resembling this is in fact what happened when the communist states of eastern Europe collapsed (see Kuran 1995). Let us call a legal system in which officials comply with the secondary rules as fragile social norms a *fragile legal system*. It is a considerable advantage of my suggestion that we conceive of officials as complying with secondary rules as social norms as Bicchieri defines them that it is able to identify and make sense of this kind of scenario.

To illustrate how latent social norms help us to identify and make sense of some of the ways reality can depart from the 'normal, unproblematic case', consider the following example. Imagine a powerful state colonises a territory half way across the world. The coloniser state absorbs the colonised territory into its legal system by appointing officials from the coloniser state to govern the territory; these officials accept and comply with the secondary rules of the coloniser state, regarding them from the internal point of view. Over time, promising locals from the colonised territory are identified and trained to become officials of the system. These trainee local officials are socialised into the practices and traditions of the coloniser state and come to endorse and identify with them. In particular, they come to accept the secondary rules of the coloniser state and, when the time comes for them to become fully fledged officials, they comply with the secondary rules of the coloniser state and regard them from the internal point of view. Eventually, all officials governing the colony are local officials of this kind.

This situation persists for a few generations, but at some stage the elites – including the officials – of the colony come to reconnect with their pre-colonial history, culture, and traditions – including legal tradition. As this cultural revival gathers pace, the local officials learn more and more about the ancient legal traditions of the territory and come to develop a deep affection for them, eventually coming to regard them from the internal point of view. These personal normative beliefs are not sufficient to motivate compliance with the secondary rules of the ancient legal tradition, however. Instead, compliance with these secondary rules is conditional on believing that most other officials comply and believe officials should comply. The local officials continue to comply with the secondary rules of the coloniser state, initially because they are not aware that the affection for the ancient legal system is widely shared. And local officials continue to comply with the colonial state's secondary rules even after they become aware of each other's personal normative beliefs, in large part because they fear the consequences of non-compliance. Instead of immediately deviating from the coloniser state's secondary rules, the local officials continue to comply while at the same time devising a plan to gain independence from the coloniser state, after which they will (re-)introduce the ancient legal system in the territory.

The purpose of this example is not to describe anti-colonial struggles as they have actually happened, but to present a simplified and stylised illustration of a scenario that resembles anti-colonial movements and is not itself implausible. In the example that I have described, the colony remains part of the legal system of the coloniser state while the local officials continue to comply with the latter's secondary rules. But at the same time there is a set of secondary rules – those of the ancient legal system – that are latent social norms amongst the officials: they would be complied with if officials believed that most other officials complied with them and believed officials should. Compliance with these social norms is very likely to emerge once the coloniser state has left the scene and local officials are free to change their behaviour. Moreover, if the ordinary citizens in the colony are likely to obey the rules deemed valid by officials complying with the secondary rules of the ancient legal system, then the emergence of compliance with these ancient secondary rules will also bring a new legal system into being. Let us call such as legal system a *latent legal system*. Again, it seems to me a considerable advantage of my suggestion that we conceive of officials as complying with secondary rules as social

norms as Bicchieri defines them that it is able to help us to see and understand this kind of scenario.

One might attempt to preserve Hart's claim that, in order for a legal system to exist, the officials of the system must regard the system's secondary rules from the internal point of view, in the following way. One might insist that, despite appearances to the contrary, the communist states of eastern Europe and perhaps also some colonies prior to independence were not in fact governed by legal systems. This implication would follow from a strict adherence to Hart's view because Hart claims that officials having personal normative beliefs that correspond to the secondary rules of the system 'is logically a necessary condition of our ability to speak of a single legal system', but in both the case of the communist states of Eastern Europe and potentially also some colonies prior to independence, the officials of what looks like a legal system do not in fact have personal normative beliefs that correspond to the secondary rules of that system. This, it seems to me, is a deeply counterintuitive conclusion to reach, and one that appears to go against Hart's contention that a society can deviate from the 'normal, unproblematic case' and still have a legal system. Instead of drawing such a conclusion, it seems to me better to make sense of such situations as legal systems that depart from the 'normal, unproblematic' case because the secondary rules of the system are either fragile or latent social norms for the officials. The great strength of Hart's account of law is its explanatory power and this amendment to it boosts it explanatory power further still.

2.4. The complexities and controversies of social justice activism

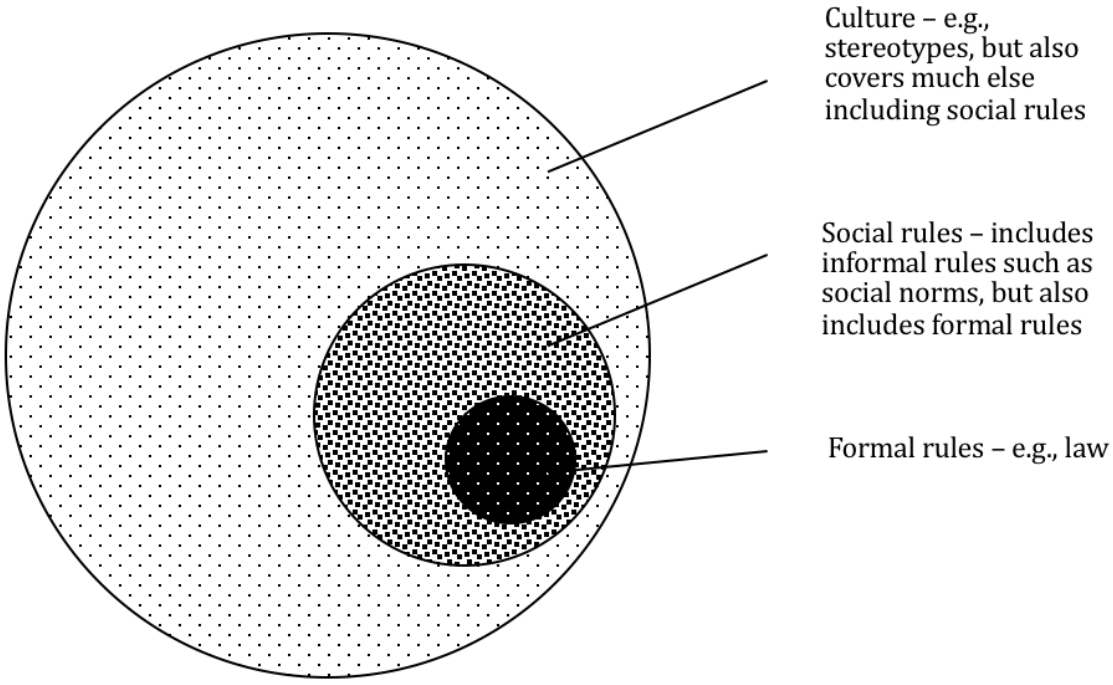
I have covered quite a lot of ground and some parts of this chapter have been quite involved, so let me summarise what I have discussed so far. I have introduced culture, informal rules, and formal rules as three categories of social phenomena, and I have focused in particular on three instantiations of each phenomena that can generate social injustice.

Culture refers to the mental and/or public representations that are widely distributed and enduring within a population. All of the social phenomena I am interested in in this chapter are features of culture, but I focused on stereotypes in particular because they are not also covered by the other two categories and I think that they can generate social injustice in a distinctive way. A stereotype is a widely distributed, simplified representation of a social group.

A social rule is a rule that is widespread and enduring in the minds of the population to which it applies, and an informal (social) rule is a social rule that is not part of a network that includes secondary rules about how the rules of that network are identified, created, modified, applied, interpreted, and enforced. The specific instantiation of an informal rule I focused on was a social norm. A social norm is a rule that people prefer to follow on the condition that they believe that most other members of their reference network: (a) follow the rule and (b) believe that they (that is, the conditional followers) should follow the rule.

Finally, formal rules are social rules that are part of a network that includes secondary rules about how the rules of that network are identified, created, modified, applied, interpreted, and enforced. The specific instantiation of a formal rule I focused on was a law. And I understand a law to be a formal rule whereby the secondary rules of the system it is part of are complied with by state officials – that is, those members of society who successfully lay claim to the monopoly of legitimate violence. Figure 2 below illustrates these various social phenomena and how they relate to each other.

Figure 2: Three categories of social phenomena and the instantiations of them that can generate social injustice



I will now very briefly sketch some of the ethical complexities and controversies associated with seeking to remedy social injustice underpinned by each of the specific mechanisms I have focused on above. Here, my aim is not to resolve these complexities and controversies but merely to outline them; they will be explored in much greater detail in the chapters to follow. In both the rest of this chapter and the rest of this dissertation, I will discuss the sites of social justice activism in reverse of the order that I have presented them so far. This is in part because we are more familiar with social justice activism that seeks to intervene in law and the ethical complexities and controversies such interventions throw up, and so it makes sense to start there. I also hope that the insights gained from this more familiar topic will help to illuminate the less well-worn terrain of social norms and stereotypes.

2.4.1. Intervening in law: radical activism and political authority

Recall that laws are rules that are part of a network that includes secondary rules that stipulate how officials are to create, modify, apply, interpret, and enforce the rules of the network. Given that such rules are formal rules, let us call any attempt to remedy social injustice by changing the law *formal activism*. We might appeal to the presence of secondary rules to distinguish between two kinds of formal activism. First, there is formal activism which does not violate the rules about how it is that laws should be changed. Let us call this kind of formal activism *ordinary activism*.[26] What counts as ordinary activism will vary depending on what the secondary rules are in the system in question, but typical examples include voting for the candidate one believes is most likely to advance social justice or attending a legal protest to demonstrate one's opposition to social injustice. Second, there is formal activism which violates the secondary rules which specify how it

[26] In calling this form of activism 'ordinary', I do not mean to suggest that it is unimportant or of lesser value. Far from it. A great deal of social injustice has been remedied via these means. It is also worth noting that it is possible to distinguish between two forms of ordinary activism. The first, which we might call *codified ordinary activism*, involves attempting to remedy social injustice by making use of participatory avenues explicitly made available by the law. The second, which we can call *uncodified ordinary activism*, involves attempting to remedy social injustice via means that, though permitted by the law, are not officially designated avenues of participation. Which activities count as codified ordinary activism and which count as uncodified will depend on the secondary rules of the legal system in question. In the UK, an example of the former is voting, while producing highly political pop music is an example of the latter.

is that laws are to be changed. Let us call any attempt to intervene in law to remedy social injustice by engaging in acts that violate the secondary rules about how laws are to be changed *radical activism*. Again, what counts as radical activism will vary depending on the secondary rules of the legal system in question, but activities that involve breaking (primary) laws or violence are highly likely to be proscribed by the secondary rules of most legal systems.[27]

Radical activism includes but is not limited to the form of social justice activism widely known as 'civil disobedience' – that is, illegal activity designed to encourage the state to change the law. Much has been written about civil disobedience, but I discuss it in the context of the broader practice of radical activism because it seems to me that, whether in the case of civil disobedience in particular or radical activism more generally, the question of justification turns on a similar set of issues. Amongst them is the fact that states claim that they have the right to rule and that their citizens have a correlative duty to obey the state's commands because the state has issued them. That is, states claim to have *political authority*.

[27] It is worth stressing that radical activism, as I understand it, is defined by the fact that it violates the *secondary rules* of the system, not the primary rules. It seems to me that secondary rules typically proscribe changing law or policy via means that are illegal (i.e., in violation of primary rules) and/or violent, but I don't think that there is any necessity to this. Secondary rules could in principle stipulate anything and so there is no conceptual impossibility about secondary rules that identify certain otherwise illegal and/or violent activities as means by which laws or governments might be changed. Imagine, for instance, a secondary rule that stipulated that someone with a claim to the crown can legitimately usurp the current monarch by defeating him a duel. It is also worth saying that I don't think that the secondary rules of the system need to be written down or codified. They may simply be social norms amongst the officials. For a recent illustration of the fact that radical political action does not necessarily involve acts that are illegal (i.e., in violation of primary rules) and/or violent, consider the UK Government's controversial attempt to prorogue Parliament in 2019. This was widely perceived as an attempt by the Government to pursue its policy agenda by ending the parliamentary session. This act did not violate any primary rule of UK law and neither was it violent. Nevertheless, the UK Supreme Court considered whether the act was constitutional – that is, in keeping with the UK's secondary rules about how government ought to be conducted and government policy ought to be pursued. The UK Supreme Court ruled that the Government's actions violated these rules and so were unlawful. The UK Government's attempt to prorogue Parliament in 2019 thus counts as a radical act, on my view.

If a state is morally justified in claiming political authority, then this seems to suggest that radical activism is impermissible because it involves citizens violating their duty to obey the law. However, if a state lacks the right to rule and citizens lack the correlative duty to obey, then radical activism is not at all controversial at the bar of political authority, though it can still be assessed in light of the principles of interpersonal morality such as the conditions for permissible defensive harm. Because ordinary activism does not violate the secondary rules about how laws are to be changed, these political authority-related issues do not apply to ordinary activism. This is why, in this dissertation, I focus on radical activism, as opposed to ordinary activism: its seems to me that the former stands in greater need of moral justification.

The political authority-related framing just sketched is how the question of the justifiability of radical activism is standardly approached (e.g., Rawls 1999a, Pasternak 2019). In chapter 3, I explore the relationship between radical activism and political authority in much more detail. More specifically, I challenge the binary account of political authority that is typically presupposed in this discussion and draw out the implications of this challenge for violent activism in particular.

2.4.2. Intervening in social norms: informal activism and civil society

We are familiar with how law can generate social injustice, but earlier I tried to show that social norms can underpin social injustice too. Social norms are informal rules – rules that are not part of a network that includes secondary rules about how rules are created, modified, applied, interpreted, and enforced. Thus, let us call any attempt to remedy social injustice by changing social norms *informal activism*.

A widely held view in political philosophy is that citizens' duties to engage in social justice activism are state mediated (e.g., Rawls 1999a). This means that justice does not require citizens to attempt to remedy social injustice directly but instead demands that we encourage the state to do so. Some argue that the fact that social injustice can be generated by social norms means that individuals' duties to engage in social justice activism need not be state mediated; it is up to those who comply with the norms to change them (Cohen 1997). Others argue, in response, that social injustice generated by social norms can in fact be remedied by state action in the form of law and policy (Ronzoni 2008, Schouten 2013). Such state action would seek to counteract the effect of unjust social norms and encourage compliance with just ones. So, returning to the example

raised earlier, the state might seek to encourage compliance with the non-abuse rule via the education system or public advertising.

The suggestion that the state should remedy social injustice generated by social norms is good as far as it goes. The problem with it is that it does not reckon with the fact that the state's ability to legitimately intervene in social norms is limited. This is because social norms are often incubated and propagated within civil society – the sphere within which citizens can, and should, act relatively free from state coercion. This presumption in favour of freedom means that unjust social norms can be incubated by and propagated within associations such as families and religious groups, and the state will be limited in how much it is able and permitted to do to successfully undermine these norms. The implication of this, I think, is that informal activism need not be state-mediated; sometimes, those who seek to remedy social injustice generated by social norms should seek to change these norms directly.

In chapter 4, I explore the relationship between informal activism, state-mediation, and civil society in much greater detail. More specifically, I identify the values that explain why the state is limited in how much it can legitimately intervene in civil society and argue that these values also inform how informal activism should be practised.

2.4.3. Intervening in stereotypes: social justice activism and social environments

Of the three social phenomena I have introduced in this chapter, it is probably stereotypes – those simplified representations of social groups that are widely distributed in a population – that operate least transparently. But, as I highlighted in section 2.1, stereotypes can generate social injustice, just as laws and social norms can. This means that, in order to remedy social injustice, it will sometimes be necessary to intervene in stereotypes. However, given that the process by which stereotypes generate social justice is rather subtle and mysterious, to intervene wisely, we must gain a better understanding of how stereotypes work. Hence, my aim in chapter 5 is to explore the distinctive role that stereotypes can play when it comes to generating social injustice, and to identify the implications of this role for the practice of social justice activism.

My strategy is to begin by examining a recent debate between Elizabeth Anderson and Tommie Shelby regarding the ethics of integration, and then to use the insights this investigation yields as a platform on which to build an account of activism that I call

stereotype activism. *Stereotype activism* involves attempting to remedy social injustice by changing the stereotypes that are present in a culture.

There are two aspects of stereotype activism I focus on in particular. The first is the extent to which stereotype activism involves intervening to change social environments, because the representations of social groups that constitute stereotypes are often embedded in the social environment. I suggest that this highlights that there are multiple means by which social justice activism might be pursued and that the relationship between these means is often complicated. The second aspect builds on the first. Because, stereotype activism can be pursued by changing the social environment, it is a form of influence that need not involve deploying sound or cogent arguments, and one might object to influence of this kind. I call this *the manipulation worry*.

2.5. Conclusion

In this chapter, I have introduced three kinds of social phenomena which can generate social injustice – stereotypes, social norms, and law – amending and integrating the latter two phenomena along the way. I called these mechanisms the three sites of social justice activism because, in the chapters to follow, I will explore how activism which intervenes in each one should be practised. In part because, of the three, it is the one with which we are most familiar, I start in the next chapter with a discussion of how to intervene in law.

3. Law, political authority, and the ethics of radical activism

> Our problem was not whether to fight, but how to continue the fight.
>
> Mandela 1964

Introduction

In the previous chapter, I distinguished between three different mechanisms that can generate or underpin social injustice, and perhaps the most widely recognised is the law and policy of the state. Examples of social injustice generated by this mechanism include laws which prohibit sexual relations between consenting adults of the same sex, laws and policies which severely limit the reproductive or economic freedoms and opportunities open to women, or laws which deny political rights to members of an unfavoured race. Recall that any effort by citizens to remedy social injustice can be called social justice activism. An important question for any theory of social justice activism is: how should social justice activism that aims to encourage the state to change its laws and policies be practised?

Laws are the paradigmatic example of formal rules: rules that are part of a system that includes secondary rules about how the rules of that system are identified, created, modified, applied, interpreted, and enforced (Brennan et al. 2012, also see Hart 2012). Hence, as I suggested in the previous chapter, we might label attempts to remedy social injustice by changing the law formal activism. Given that one of the characteristic features of legal systems is that they include rules about how the laws of the system are to be changed, one useful distinction to draw is between those forms of formal activism that do not violate the systems' secondary rules about how laws change and those that do.

Once again, as previously suggested, let us call the former ordinary activism. What counts as ordinary activism will of course vary depending on what the secondary rules in the system in question are, but examples include: gradually changing public opinion in the hope that this will lead citizens to pressure their state to change the law; persuading state officials that the status quo is in fact inconsistent with the basic, fundamental rules of the system; or organising mass strikes or protests in order to demonstrate the extent of public opposition to the current arrangement. To call this kind of activism 'ordinary' is in no way intended to downplay its value or importance. A great deal of social injustice

has been remedied via these means. Nevertheless, ordinary activism is limited, for at least two reasons.

First, there may be situations in which the secondary rules of the system drastically limit the opportunities available for citizens to contribute to changing the law. In which case, very little will count as ordinary activism. Second, even if the secondary rules of the system allow citizens to contribute to legal change via a range of avenues, there is no guarantee that influence via these means will always be successful. Those who hope to influence the state to remedy social injustice therefore have one option remaining: they can engage in activism that violates the state's secondary rules about how laws are to be changed. This is what I call radical activism.

What counts as radical activism will again vary, but, as I mentioned in the previous chapter, I take it that activities that involve breaking (primary) laws or violence are highly likely to be proscribed by the secondary rules of most states. Henceforth, when I speak of radical activism I will be referring to attempts to influence the state to change the law by means that are either illegal or violent. This chapter explores the ethics of such practices.[28]

This topic is often considered in light of the fact that states claim to have political authority, which is taken to consist of a state's right to rule and its citizens' correlative duties to comply with the state's commands. Citizens' duties to comply with the state's commands are said to be *content-independent*, which means that they do not depend on the content of what is being commanded (Hart 1982:243-268, also see Raz 1986:35-7). Put differently, citizens are to obey the state's commands, not because of any specific virtues any particular command may possess, but *because they are the state's commands*. Framing the issue in this way helps to highlight at least some of what seems to make radical activism controversial. If states are morally justified in professing their right to rule and their citizens' duties to obey – that is, if states have *legitimate* political authority

[28] I take the term 'violence' to refer to the "deliberate exercise of physical force against a person, property, etc" (OED 2021). Some authors use 'violence' to refer to a broader range of cases including rights violations and psychological harm (e.g., Morreal 1991). While I recognise the rhetorical potency of such an approach, I prefer to use 'violence' to refer to a narrower class of cases. There are many different ways in which human life can be made to go less well and one particularly interesting and significant way is for it to be intruded on by deliberate physical force. Even if we broadened the concept of violence to cover a wider range of undesirable actions and events, we would still need a way of referring to this distinct, narrower class.

– then the burden is on the advocates of radical activism to explain how practices that fail to comply with the rules of the state can be permissible.[29]

Some of those who consider the ethics of radical activism in the light of political authority have an account of political authority that is both binary and grounded in the value of justice. An account of political authority is *binary* if it suggests that there are only two statuses that a regime can have with respect to authority: either it has full authority, or it has no authority. An account is *justice-grounded* if it appeals to the value of justice to explain the source of the state's authority. Putting the two together, we can say that a binary, justice-grounded account of political authority is one according to which states are either at least nearly just and so have full authority or are seriously unjust and so lack all authority.

Considering the ethics of radical activism in the light of a binary, justice-grounded account of political authority yields the following analysis. When a state is nearly just and so has full authority, radical activism is highly controversial at the bar of authority. This means that authority imposes demanding constraints on the practice of radical activism; citizens are either not permitted to engage in radical activism or, in order to be so permitted, they must go to great lengths to demonstrate their respect for the authority of their regime. However, when a state is seriously unjust and so lacks all authority, radical activism is not at all controversial at the bar of authority. Authority imposes no constraints on the practice of radical activism. Instead, the permissibility of any particular instance of radical activism is determined by appealing to principles of interpersonal morality, such as the conditions for permissible defensive harm, and authority plays no part in such considerations. This is how John Rawls frames his canonical treatment of a form of radical activism he calls 'civil disobedience', and, more recently, how Avia Pasternak approaches the question of the justifiability of a form radical activism she calls 'political rioting'.

[29] Some scholars use the label 'legitimacy' to denote what I mean by 'authority'. I prefer the latter term because I think it is important to distinguish the genuine acceptance of the state's claims to have the right to rule amongst citizens, on the one hand, from the morally justified claim to have the right to rule, on the other. It seems appropriate to use the modifiers 'sociological' to pick out the former, and 'legitimate' to pick out the latter. Given this, using the label 'legitimacy' to denote the right to rule generates the slightly awkward and potentially confusing locution 'legitimate legitimacy', so I prefer the term authority instead. Nothing of substance, as far as I can tell, turns on this terminological decision.

Let us call this analysis of the relationship between radical activism and political authority the *binary, justice-grounded view*. While I agree that radical activism should be considered in the light of political authority, my aim in this chapter is to disaggregate the account of political authority that the binary, justice-grounded view presupposes, and to highlight an implication of this disaggregation for the practice of radical activism.[30] In particular, I argue that the binary, justice-grounded view ignores a crucial authority relation that can obtain in conditions of serious injustice, and that, when this relation obtains, the permissibility of radical activism is more constrained than the binary, justice-grounded view suggests.

This chapter proceeds as follows. In section 3.1, I present Rawls's and Pasternak's treatments of the ethics of radical activism as paradigmatic examples of the binary, justice-grounded view. The next two sections disaggregate the account of political authority presupposed by the binary, justice-grounded view. In section 3.2, I invoke the Hohfeldian analysis of rights to show that the right to rule is not binary but can come in many forms, a number of which have been advanced in the literature. According to one of these forms, the fact that a state possesses political authority does not entail that its citizens' have content-independent duties to obey the law. I call this kind of political authority *weak authority*. In section 3.3, I argue that order and security can explain the source of weak authority, thereby showing that values other than justice can ground the right to rule. In section 3.4, I draw out an implication of this disaggregation for the practice of radical activism. In particular, I argue that weak authority imposes more demanding constraints on the practice of radical activism than are present in the absence of all authority. This conclusion matters because there are many societies where radical activism may be called for which are governed by states that possess weak authority.

[30] Robert Jubb has recently offered a somewhat similar critique of binary accounts of political authority. Jubb's strategy is to argue that a binary account of political authority generates deeply counterintuitive judgements about the permissibility of different forms of radical activism that have taken place within different regimes. Jubb then suggests that this result compels us to disaggregate political authority, though he says very little to say about what such a disaggregation might involve (Jubb 2019). While I find Jubb's argument highly suggestive, I have doubts about the strategy he employs to reach his conclusion. This chapter therefore takes a very different approach. Instead of appealing to intuitions in an effort to persuade that the concept of authority should be disaggregated, I think a better route is to proceed more directly by demonstrating how the concept can be disaggregated and highlighting the implications of this effort.

Moreover, I also suggest that the account of the relationship between authority and radical activism that I develop can illuminate a plausible normative rationale for the approach that the African National Congress took to resisting apartheid in South Africa. Section 3.5 concludes.

One point of clarification before I proceed. In this chapter, I will take it as given that regimes that are at least nearly just have the right to rule correlated with the duty to obey the law as such held by citizens. This justice-based account of the grounds of (full) political authority has been criticised but I can safely assume it here because my interlocutors take it to be correct.[31] My disagreement with proponents of the binary, justice-grounded view is not that they are wrong to think that justice grounds the state's right to rule. Rather, my claim is that they are wrong to assume that there are no other authority relations – grounded in other values – that can exist between a state and its citizens. Moreover, I claim that at least one of these other authority relations has an important implication for the ethics of radical activism.

3.1. The binary, justice-grounded view

The binary, justice-grounded view considers the ethics of radical activism in the light of an account of political authority according to which regimes either have full authority if they are nearly just or lack all authority if they are not. According to this definition, John Rawls's theory of a particularly respectful form of radical activism, which he calls 'civil disobedience', is a paradigmatic example of the binary, justice-grounded view.

[31] A. John Simmons offers perhaps the most powerful critique of accounts of political authority grounded in the value of justice (Simmons 1979). According to a very popular version of such accounts, we have duties to comply with the laws of a just state so long as the institutions of that state "apply to us" (Rawls 1999a:294). Simmons argues that the only plausible reading of the requirement that state institutions 'apply to us' is that we have engaged in some voluntary act binding ourselves to a state. However, very few people have in fact engaged in such acts, so any given state will have political authority over very few people on this view. On the other hand, if we remove the requirement that state institutions 'apply to us' then we will not be able to make sense of the fact that states claim authority over *particular* people and citizens have duties to obey the laws of *particular* states. I cannot fully address this critique here, but I believe that the best response involves interpreting the 'apply to us' requirement along associative lines. This means that just states have full political authority over those citizens who identify themselves with the practices and values of the state (see Renzo 2012).

Civil disobedience, as it is popularly understood, is a form of radical activism because it involves citizens engaging in illegal activity in an attempt to encourage the state to change the law. However, civil disobedience, as defined by Rawls, is also a particularly respectful form of radical activism because of four further features Rawls attributes to it: first, it is nonviolent, because those who practise it wish to remain "within the limits of fidelity to the law" (Rawls 1999a:322); second, it "invokes the commonly shared conception of justice that underlies the political order"; third, it is "engaged in openly and with fair notice" (Rawls 1999a:321); and, fourth, those who engage in it "accept the legal consequences of [their] conduct" (Rawls 1999a:322).

These features are explained by the fact that Rawls's theory of civil disobedience "is designed only for the special case of a nearly just society, one that is well-ordered for the most part but in which serious violations of justice nevertheless occur" (Rawls 1999a:319). Rawls regards the state in such a society as having authority, grounded in the fact that it has a constitution which is an imperfect political procedure designed to realise social justice. This constitution "is imperfect because there is no feasible political process which guarantees that the laws enacted in accordance with it will be just" (Rawls 1999a:311). Rawls assumes that a just constitution will be democratic and involve majority rule, and so a key source of injustice will be the fact that "[m]ajorities (or coalitions of minorities) are bound to make mistakes" (Rawls 1999a:311). Citizens should nevertheless comply with the unjust laws enacted by such a constitution because we have a "natural duty to uphold just institutions" (Rawls 1999a:311).

Thus, for Rawls, justice grounds a duty to obey the law that does not depend on the virtues any particular law may possess, but on the fact that it emanates from a just though inevitably imperfect procedure. However, the fact that a just constitution can enact unjust laws means that citizens may face a "conflict" between "the duty to comply with laws enacted by a legislative majority" and the "duty to oppose injustice" (Rawls 1999a:319). The respectful features of Rawls's account of civil disobedience noted above are how he thinks citizens should acknowledge their states' authority despite breaking the law.

It is worth noting that, though some theorists regard the duty to obey the law as exhausting citizens' political obligations,[32] this does not appear to be Rawls's view. Instead, Rawls argues that, when injustice occurs in society, citizens have duties to remedy it. This duty is grounded in the same principle that generates the duty to obey the law in nearly just societies: the "natural duty… to comply with and do our share in just institutions when they exist and apply to us" and "to assist in the establishment of just arrangements when they do not exist, at least when this can be done with little cost to ourselves" (Rawls 1999a:294).[33] For the reasons set out earlier, citizens who seek to remedy injustice may sometimes need to engage in radical activism, such as civil disobedience. In chapter 1, I argued that we should reject Rawls's 'little cost' proviso, because the costliness of citizens' duties to engage in social justice activism will vary across the population and may be quite high for some. Nevertheless, I also argued that there should be a cap on the level of cost that an individual can be required to bear to discharge their remedial duties. While the burdens associated with many forms of radical activism are likely to exceed the cost cap, the value of justice will still presumably generate a reason – though not a duty – for that citizen to engage in that action.

[32] George Klosko, for instance, announces on page 1 of his book titled *Political Obligations* that he will "treat the question of political obligation as basically interchangeable with why people should obey the law" (Klosko 2005:1 [note 1]). Similarly, in the first sentence of their *Stanford Encyclopedia of Philosophy* entry on 'Political Obligation' Richard Dagger and David Lefkowitz declare that they will follow "the traditional practice of equating political obligation with the moral duty to obey the law of one's country or state" (Dagger and Lefkowitz 2021).

[33] Rawls also argues that, over and above the duties grounded in the natural duty of justice, those citizens who assume political office have additional political obligations grounded in the principle of fairness, and so these citizens are are "bound even more tightly" than other citizens (Rawls 1999a:303). Rawls conceives of these additional political obligations as duties to comply with and support just institutions, and he seems to assume that the principle of fairness simply fails to generate any duties when the institution in question is sufficiently unjust. However, Candice Delmas argues persuasively that, when an institution is unjust, the principle of fairness requires those who benefit from that institution to remedy it (Delmas 2014). If this is right, then the principle of fairness mirrors the natural duty of justice, in that it generates duties to comply with and support institutions when those institutions are just, and it generates duties to remedy institutions when those duties are unjust. The constituency amongst whom these duties are generated is different of course. In the case of the natural duty of justice, it is all those to whom the institution applies. In the case of the principle of fairness, it is all those who accept the benefits of the institution.

Rawls limits his theory of civil disobedience to nearly just societies because he believes that activism that has the features he describes does not arise as a problem in societies that are not at least nearly just. In societies that are seriously unjust, the state has no right to rule and citizens have no duties to obey its commands simply because they are its commands. Rawls describes such societies as "a kind of extortion, even violence" (Rawls 1999a:302) and asserts that civil disobedience presents "no difficulty" in these cases (Rawls 1999a:319). Indeed, he even suggests that violence may be justified in order to remedy injustice in such societies (Rawls 1999a:310, 319, 323, 328).

Avia Pasternak has recently provided another very clear example of the binary, justice-grounded view. Pasternak is interested in a particular form of radical activism she calls political rioting, which involves citizens "resort[ing] to spontaneous, disorganised, public collective violence in order to protest against and to defy their political order" (2019:385). Moreover, her focus is on political riots that respond to the kinds of social injustices that characterise contemporary democracies such as France, the United Kingdom, and, especially the USA. These are societies in which, "despite overall levels of prosperity, portions of their populations, and often racial minorities in particular, experience serious material deprivation, social exclusion, and political marginalisation" (Pasternak 2019:389). Pasternak regards these societies as seriously unjust and conceives of political rioting as a means by which the citizens of such societies seek to remedy the social injustice they face. After considering whether political rioting is compatible with states' claims to political authority in the societies she focuses on, Pasternak concludes that citizens are not under a duty to respect the authority of the state in societies marred by serious injustice (Pasternak 2019:394-8). The clear implication of her view is that, in societies that are not at least nearly just, the state lacks all authority and so political rioting is not at all controversial at the bar of authority.

This does not mean that the political rioting is straightforwardly permissible, however. It involves violence and the imposition of harm and so, in order to be justified, it must meet the conditions of permissible defensive harm. Pasternak invokes three such conditions.[34] First, there is the success condition, which requires that the use of defensive

[34] Civil disobedience may also involve the imposition of harm and Rawls invokes similar conditions for civil disobedience to be justified (Rawls 1999a:326-331). These conditions apply in addition to the respectful features Rawls builds into the definition of civil disobedience.

harm has a reasonable prospect of averting or at least ameliorating the attack that triggered it. Second, there is the necessity condition, according to which defensive harm is permissible "only if it risks imposing the least morally weighted harm of all the available alternatives that can be deployed in order to deter an unjustified attack" (Pasternak 2019:401). Third, there is the proportionality condition, which requires a calculation in which the expected magnitude of the benefits that defensive action will deliver must be weighed against the expected magnitude of the costs that the action will involve. In this calculation, the appropriate ratio of benefit to harm varies depending on the culpability of the agent being targeted. Agents who culpably expose victims to unjust harm are liable to a roughly equal level of defensive harm. However, when it comes to other agents in society – those that are not culpable for injustice – imposing harms on them can only be justified "if the benefits yielded by imposing them are very substantial in comparison to the harm they cause" (Pasternak 2019:406-7).

Because Pasternak regards the societies she focuses on as seriously unjust and so totally lacking in authority, authority imposes no constraints on the practice of radical activism. The justifiability of political rioting is determined by appealing to the conditions for permissible defensive harm, and political authority plays no part in these considerations.

3.2. The right to rule is not binary

The intuitive idea underlying political authority is *the right to rule.* The binary, justice-grounded view presupposes an account of political authority according to which the state either has the right to rule which is necessarily correlated with citizens' content-independent duties to obey, or the state does not have the right to rule and so citizens do not have content-independent duties to obey. Having the right to rule is therefore taken to be a binary, all-or-nothing affair. The implication of this is that radical activism can only be either highly controversial at the bar of authority (i.e., when it is present) or not at all controversial at the bar of authority (i.e., when it is absent).

The problem with the account of political authority presupposed by the binary, justice grounded view is that it assumes that there is just one conception of the right to rule, such that either states have the right to rule or they don't. However, there are, in fact, multiple conceptions of the right to rule, which means that the question of whether a state has the right to rule or not does not admit a simple, binary, yes-or-no answer. To

see how it is that there are multiple conceptions of the right to rule we must first see that there are multiple conceptions of rights. According to Wesley Hohfeld's widely accepted taxonomy of rights, rights can be divided into four categories or 'incidents'. First, there are privilege-rights which entail having no duty not to do something. Second, there are claim-rights which entail that others have duties to do or not do something, where these duties are owed to the right holder. Third, there are power-rights which entail having the ability to alter one's own or another's rights. And fourth, there are immunity-rights the having of which entails that others lack the ability (i.e., the power-right) to alter one's rights (Hohfeld 1923:36-64). Different conceptions of the right to rule can be understood by appealing to one or more of these Hohfeldian incidents.[35] I will highlight three.

According to the first conception, having political authority involves having a power-right to place citizens under genuine obligations by issuing directives about how citizens are to act, but these obligations are not necessarily owed to the state (e.g., Raz 1986:23-105).[36] According to the second conception, the state's right to rule is a cluster of Hohfeldian incidents, including a power-right to impose duties on citizens and a claim-right to the obedience of citizens (e.g., Simmons 2001:130).[37] Unlike the first conception, in this case, citizens' duties to obey are owed to the state. These two conceptions of the right to rule are perhaps the most prominent in the literature, but we can also appeal to Hohfeldian incidents to identify a third. According to this third conception, the state's right to rule is a privilege-right. While the first two conceptions entail content-independent duties to obey held by citizens, this conception does not. On this view, to have political authority is to be justified in issuing the commands and engaging in the coercion that ruling necessarily involves (e.g., Ladenson 1980, Wellman 1996).

[35] Judith Jarvis Thompson refers to rights comprised of more than one Hohfeldian incident as "cluster-rights" (Thompson 1990:55).

[36] Raz, for instance, argues that the state serves citizens by helping them to comply with reasons that were already operative for them – state directives are reasons for action that are based on but replace the reasons that independently apply to those to whom the directives are issued.

[37] I take it that this conception is constituted by a power-right and a claim-right, and not a claim-right alone, because the state's ability to change others' rights – which includes but is not limited to the ability to impose genuine obligations – is more fundamental than whether the obligations the state imposes are owed to the state or not. Many of the important functions of the state – conferring power on judges, for instance – do not involve issuing commands to citizens.

So, we have at least three conceptions of the state's right to rule: as the power-right to place citizens under binding obligations; as the power-right to rule plus the claim-right to the obedience of citizens; and as the privilege-right to rule which does not entail any correlative duties. Proponents of these different accounts can sometimes speak as though their preferred conception is the 'correct' one (e.g., Raz 1986:24-8). However, once we see that we can appeal to Hohfeldian incidents to cash out the right to rule in different ways, there is no need to choose between these three conceptions or to insist that one is ultimately 'correct'.[38] As Thomas Christiano says after outlining a similar taxonomy, "[i]t is not a useful aim of philosophers or political thinkers to determine which one of these conceptual accounts of political authority is the right one. Each one of them grasps a kind of legitimacy of political authority that is worth taking into account and distinguishing from the others" (Christiano 2012).

Moreover, while it is not entirely clear which conception of the right to rule proponents of the binary, justice-grounded view have in mind, it is important to note that their discussions presuppose one of the first two just listed – that is, a conception of the right to rule that entails citizens having duties to obey the state's commands simply because they are the state's commands. Ignoring the differences between the two conceptions, let us say that a state that has either the power-right or the power-right plus the claim-right to rule has *full authority*. The fact that there is a valid conception of the right to rule that does not entail citizens' duties to obey suggests that a position between a state having no authority and a state having full authority is possible. Call this *weak authority*.[39]

---

[38] I don't want to suggest that these three conceptions of the right to rule exhaust all the plausible possibilities.

[39] Some have labelled what I am calling weak authority 'legitimacy' and reserved the label 'authority' only for what I am calling full authority (e.g., Buchanan 2002, Huemer 2013:5). I have not taken this approach for two reasons. First, as I mentioned in note 29 above, I think it is important to be able to mark the distinction between a state's claims to have the right to rule being generally accepted by its citizens and a state having the morally justified right to rule. I call the former *sociological* political authority and the latter *legitimate* political authority. But given that it is perfectly possible to draw this distinction with respect to the privilege-right to rule, labelling it 'legitimacy' generates awkward and potentially confusing locutions (e.g. 'legitimate legitimacy'). The second reason for favouring the terminology I propose is that both weak and full authority can plausibly be seen as instantiations of the right to rule. It therefore makes sense to regard them as conceptions of that broader concept, and to label them likewise.

My aim in this section was to show that the right to rule is not a binary, all-or-nothing affair and I take it that the possibility of weak authority demonstrates that there are in fact multiple conceptions of the right to rule. In the next section, I turn to discuss the grounds of these different conceptions.

3.3. Values other than justice can ground the right to rule

The ground of political authority is the value (or values) that is invoked to explain or justify the state's right to rule. As we have just seen, conceptions of the right to rule can be divided into those that entail citizens' duties to obey the law, on the one hand, and those that do not entail citizens' duties to obey the law, on the other. Proponents of the binary, justice-grounded view presuppose the former conception of political authority and, moreover, ground political authority in the value of justice. For them, a state has full political authority (which entails citizens' duties to obey the law as such) if and only if it realises a nearly just society via laws enacted by an inevitably imperfect democratic constitution. As I said in the introduction of this chapter, I will simply assume that justice does indeed ground full authority. My aim in this section is to show that we can invoke a different value to ground weak authority.

A number of principles have been appealed to in order to justify the state's privilege-right to rule (see, e.g., Ladenson 1980, Wellman 1996, Buchanan 2002), but my own preference is to ground weak authority in the value of order and security. The broad contours of the Hobbes-inspired account I favour are very familiar, but some of the details are distinctive so let me briefly sketch it here. We would start from the observation that, when large numbers of human beings who do not share ties of kinship or close community live in close proximity to one another in the state of nature, they tend to be drawn into mutually destructive conflict which makes life intolerably insecure. The reason for this is not that all would constantly seek to predate on others, but that some would, and one cannot know in advance who they are. Individuals threaten each other in the state of nature, therefore, in the sense that their proximity to each other generates this intolerable insecurity.

If we assume that all individuals have a natural duty not to pose threats to others (Renzo 2011), then we can say that all individuals have a duty to end the conditions that characterise the state of nature. The problem is that individuals cannot end these conditions on their own. Even if they could agree on common rules of co-existence, they

lack the power to enforce them and so any initial compliance would quickly unravel. The state solves this problem by using the threat of overwhelming force to give those who would seek to violate common rules a compelling incentive to comply and those who would not seek to violate common rules assurance that their compliance will not be exploited.

The state, therefore, is the means by which we discharge our natural duty not to pose threats to one another, and the state's privilege-right to rule is grounded in it successfully fulfilling this role. According to this account, the state must provide order for *all* people living in its territory, not just some; those who are terrorised or left insecure are effectively still in the state of nature and those who are secure are posing a threat to them. This explains why regimes such as Nazi Germany or the antebellum USA lacked even the privilege-right to rule. The natural duty not to pose a threat to others does not generate a content-independent duty to obey the law, because the state can maintain order by commands and coercion alone. However, it does generate a duty to avoid actions that risk re-entry into the state of nature, once a state has been created and is successfully maintaining order.[40]

We can combine this discussion with the analysis of the concept of the right to rule in the previous section. Full authority refers to any conception of the right to rule that entails citizens' duties to obey the law as such, and a state has full authority when it realises a nearly just society via laws enacted by a democratic constitution. While political philosophers have advanced many competing accounts of ideal social justice, many would agree that few if any societies qualify as 'nearly just'. It follows from this that perhaps no regime that currently exists – or, most likely, has ever existed – has full authority.

[40] Massimo Renzo also grounds political authority in the natural duty not to pose threats, so it may be useful to compare our accounts. Renzo and I agree that merely living in close proximity to others in the state of nature constitutes posing a threat to them. Nevertheless, my view differs from his in at least two ways. First, Renzo appeals to the natural duty not to pose threats to ground the state's right to rule, conceived of as the power to create moral obligations and the claim-right to have these duties complied with, while I focus only on the privilege-right to rule which entails no duties to obey. Second, Renzo stipulates that the natural duty not to pose threats can only ground the authority of states that are "reasonably just" (Renzo 2011:578 [note 13]). I, on the other hand, argue all states that realise order and security for all and so enable individuals to discharge their duties not to pose threats have the privilege-right to rule, even unjust ones.

Weak authority refers to the privilege-right to rule which does not entail any duties to obey the law as such held by citizens, and a state has weak authority when it realises order and security for its citizens. This covers a very wide variety of regimes, including contemporary states ranging from Canada to Chile, from Spain to Singapore. However, given the attention it garners within contemporary political philosophy, perhaps the most striking example of a state that has weak authority in my view is the regime that governs the USA. The USA is widely and correctly regarded to be a seriously unjust society, and the injustice that characterises the USA is especially pronounced with respect to position of African Americans (see, e.g., Anderson 2010, Shelby 2016). The nature of this injustice is wide-ranging and pervasive, but illustrative evidence includes the fact that African Americans are massively over-represented amongst the prison population of the USA (Sawyer and Wagner 2020) and the fact that they are three times more likely to be killed by police than white people are (Buehler 2016).

I take it to follow from the fact that the USA is seriously unjust that the citizens of the USA lack content-independent duties to comply with the commands issued by their state. Nevertheless, it seems to me that the state in the USA has managed to secure minimal order and security in the territory and so has weak authority. A critic might deny that the USA has in fact managed to realise order and security for African Americans, perhaps citing as evidence the fact that African Americans experience higher rates of imprisonment and face a greater risk of death at the hands of police. But facing a greater risk of harm – even death – compared to others is not the same as experiencing the conflict and insecurity characteristic of the state of nature, and it seems to me that, on any plausible account of what a state of nature-like scenario involves, the USA has escaped it. Denying this generates the striking – and it seems to me implausible – implication that, with respect to order and security, there is currently no qualitative difference between the position of (some) blacks in the USA and the average citizen of Yemen.

There are two additional points worth bearing in mind in this connection. The first is that we should draw a distinction between two things. On the one hand, there is the objective fact of whether a state has the capacity and inclination to deploy the overwhelming force at its disposal to ensure that its citizens escape from the conditions of conflict and insecurity that characterise the state of nature. On the other hand, there is the question of whether some subset of citizens have confidence in the state's willingness to protect them, or whether citizens are themselves willing to call on the state to protect

them. One of the legacies of a history marred by horrendous state-inflicted social injustice and indeed a refusal by the state to secure order for a population may be that the population in question lacks confidence in the state's willingness to protect them, even when the state is so inclined. Another legacy may be that the population in question develops social norms which prohibit calling on the state for protection. It is likely that both legacies are present in some African American communities in the USA, but it does not follow from this that – as an objective matter – the state has failed to realise order and security. The second point to note is that two countries can both realise order and security and so both have weak authority, but also be vastly different in other morally relevant ways. So, it seems to me that the states in both Canada and the USA have realised order and security in their territories, but, while both societies are unjust, the former is less unjust than the latter.

When a state fails to realise order and security for all, it has no authority – that is, it lacks any right to rule. While I will not further specify what constitutes order and security for all, we can be sure that regimes such as the Libyan Government of National Accord, which currently claims authority in Libya but is not able to secure peaceful relations in the territory, or the Nazi regime in Germany, which claimed authority but terrorised some of its own citizens and so kept them in a state of nature-style situation, do not provide it.

3.4. Radical activism in conditions of weak authority

So far, I have disaggregated the account of political authority presupposed by the binary, justice-grounded view, arguing that (a) the right to rule is not binary but can instead take a variety of different forms, and (b) at least one of these forms – the privilege-right to rule – is grounded, not in the value of justice, but in the value of order and security. This suggests that the binary, justice-grounded view is mistaken about the relationship that obtains between political authority and citizens' duties. The binary, justice-grounded view assumes that political authority is correlated with the duty to obey the law, but the fact that a state has weak authority does not entail that its citizens have content-independent duties to obey the state's commands. It is, however, associated with citizens having duties to avoid actions that risk re-entry into the state of nature.

It is important that I am clear about the difference between the relationship of (conceptual) entailment on the one hand, and the relationship of (normative) association

on the other. Weak authority *does not entail* citizens' duties to obey the law as such because weak authority refers to the privilege-right to rule. Unlike full authority, which encompasses both the claim-right and the power-right plus claim-right to rule, it is not part of the definition of privilege-rights that they are correlated with duties. Weak authority is *associated with* citizens' duties to avoid acts which risk re-entry into the state of nature, however, because the value that grounds the former – order and security – also generates the latter. This means that, while weak authority and the duty to avoid acts that risk re-entry into the state of nature are not conceptually correlated, they do, as a matter of normative grounding, coincide.[41]

I now want to draw out an implication of the disaggregation I conducted in sections 3.2-3.3 for the practice of radical activism in conditions of weak authority. In particular, I will argue that conditions of weak authority impose constraints on the practice of radical activism that are not present when a regime is totally lacking in authority. The first step of this argument involves reflecting on the constraints that are thought to apply to radical activism in conditions of no authority. Recall that these constraints are imposed by the conditions for permissible defensive harm – success, necessity, and proportionality.

There are two features of these constraints that I think are worth noting. First, any attempt by citizens who wish to remedy social injustice to make use of the success, necessity, and proportionality conditions to guide their action is likely to be fraught with uncertainty and unpredictability.[42] Even if a citizen had a clear sense of the prospect of success that warranted the label 'reasonable', it would be empirically very difficult to be

[41] Thanks to Cécile Laborde for pressing me to be clearer on this point.

[42] Pasternak frames her discussion as a moral assessment of political rioting, suggesting an impartial, third-party perspective, but a number of remarks she makes suggest that she also intends to provide an action-guiding account for situated agents. She advertises the fact that her account does not "make demands that rioters could not possibly meet" (Pasternak 2019:387) as an advantage of her view, for instance. Pasternak also endorses evidence-based standards of the success and necessity conditions, which require that assessments of them be based on "the conclusion that a rational and unbiased person would draw from the evidence reasonably available to her" (Pasternak 2019:401), and argues that the deliberations of ordinary citizens can meet such a standard. Moreover, the various limits of permissible defensive harm that Pasternak identifies seem to be addressed directly to the citizens to be constrained. In any case, I presume that an account of the ethics of activism would be radically deficient if it could not and in no way intended to be action-guiding.

sure that the action one was contemplating has such a prospect of success. The necessity condition introduces even more challenging empirical judgements, requiring citizens to assemble all the available alternatives that have a reasonable prospect of averting the unjust threat and then to identify the option amongst these alternatives that risks imposing the least harm. The proportionality calculation may be the most demanding of them all. Here, citizens are asked to ascertain the expected magnitude of the benefits and harms associated with the actions they are considering and to determine the culpability for injustice of various agents in society, in order to assess whether the harm that an action is likely to involve can be directed at agents in society in such a way that is proportionate with the benefits the action is expected to deliver.

My point here is not that the conditions for permissible defensive harm can never be met. Rather, my point is that it will very rarely, if ever, be the case that an agent using these conditions to guide their action can be 100% (or at any rate very) confident that they have been met, especially when the action being guided is an instance of social justice activism. This means that, if the conditions for permissible defensive harm are to be operationalizable, action-guiding standards, then we must recognise that certainty is not possible (see Finlay 2010:308). Instead, citizens contemplating radical activism should be satisfied that harm is permissible when they are *sufficiently confident* that the success, necessity, and proportionality conditions have been met. The idea of being 'sufficiently confident' is of course extremely vague and, though I will have more to say about it in a moment, I will not attempt to specify it in this chapter. The key point for present purposes is that, however we specify 'sufficiently confident', it must refer to a level of confidence that is considerably lower than 100% – that is, it must refer to something less than full certainty and predictability.

The second feature worth noting is that, while Pasternak's focus is on *political rioting* in seriously unjust *democracies*, the constraints of the permissible defensive harm framework apply to other forms of radical activism in other unjust societies. Forms of radical activism vary with respect to the level of harm and disruption they are likely to cause, ranging from civil disobedience as Rawls defines it at the milder end, to insurgency and civil war at the more extreme. Unjust societies also vary with respect to the level of harm they impose on their victims, ranging from those characterised by 'mere' socioeconomic injustice, such as the contemporary democracies Pasternak discusses, to those that terrorise or enslave individuals in society, such as the regimes in Nazi Germany

or the antebellum USA. Given the level of harm associated with egregiously unjust regimes, and therefore the benefit that is yielded by remedying the pervasive injustices characterising such societies, it is not implausible to think that extremely disruptive forms of radical activism which themselves involve a great deal of harm – such as civil war – can sometimes meet the conditions for permissible defensive harm (see, e.g., Buchanan 2013).

Recognising these two features leads to a number of important insights. Consider the fact that invoking the conditions for permissible defensive harm is likely to be fraught with uncertainty and unpredictability. One site of unpredictability is the prospect of success of the resort to defensive harm, and recognising uncertainty in this area reminds us that even permissible defensive harm can fail to achieve its end. Now consider the fact that highly disruptive radical activism such as insurgency and civil war can sometimes be permissible. It is crucial to note that these campaigns typically involve attempts to frustrate and undermine the state's ability to govern and maintain order. Sometimes this is part of a wider objective of overthrowing the regime and replacing it with a new one – that is, of revolution – but this need not always be so. The frustration and undermining of the state's ability to govern and maintain order may instead be used as a tactic to encourage the regime to change laws or policies, but not as part of a wider strategy to overthrow the government.

Putting these two insights together generates the following conjunction of possibilities: there are sometimes circumstances that are so egregiously unjust that it can be permissible for citizens to resort to highly disruptive radical activism in order to remedy them; such activism must have a reasonable prospect of success to be permissible, but it can meet this condition and still fail because success is never certain; and despite failing to achieve its end, such activism may nevertheless successfully frustrate and undermine the state's ability to govern and maintain order, thereby risking the intolerable conditions of insecurity associated with the state of nature.

Allen Buchanan speaks to this line of thought when he says that "there are many ways a revolution can go wrong, so the possible outcomes are not limited to a continuation of the status quo of oppression or a *successful* revolution" (Buchanan 2013:301 italics in original). Similarly, Hannah Arendt echoes this concern when she declares: "[a]ction is irreversible, and a return to the status quo in case of defeat is always unlikely. The practice of violence, like all action, changes the world, but the most probable

change is to a more violent world" (Arendt 1970:80). There is empirical evidence that this worry is well-founded. Analysing a data set that includes 323 violent and nonviolent campaigns that took place between 1900 and 2006 and sought to overhaul existing regimes, Erica Chenoweth and Maria J. Stephan find that countries in which violent campaigns have occurred are significantly more likely to experience a recurrence of civil war, compared to countries in which nonviolent campaigns have occurred (Chenoweth and Stephan 2011:216-8).

We are now in a position to see why weak authority might impose some constraints on the practice of radical activism. A state's weak authority, recall, is grounded in its ability to realise order and security for all individuals in a given territory, thereby enabling these individuals to avoid posing the threats to one another that create the pervasive insecurity of the state of nature. When the state is successfully maintaining order and security, the natural duty not to pose a threat to others has been discharged and so demands no positive action, though it does require citizens to avoid actions that risk re-entry into the state of nature. However, a regime can maintain order and be seriously unjust; that is, a state can have weak authority and lack full authority. When this is the case, citizens have obligations/reason to remedy social injustice, and this may take the form of radical activism.

Moreover, if the regime is sufficiently unjust and the state sufficiently committed to upholding injustice, it may be permissible to discharge these obligations or act on these reasons by resorting to highly disruptive radical activism such as insurgency or civil war. Such activism may fail to remedy social injustice but nevertheless undermine the state's ability to govern and maintain order, thereby risking re-entry into the state of nature. Thus, we have the possibility of a conflict between the duty/reason to remedy social injustice – grounded in the natural duty of justice – and the duty to avoid acts that risk re-entry into the state of nature – grounded in the natural duty not to pose threats. It is this possibility that explains why the presence of weak authority imposes constraints on how radical activism should be practised.

It is worth noting that this conflict between duties is analogous to the conflict that occurs in conditions of full authority and creates the problem of civil disobedience, at least as Rawls understands it. Recall that the state in a nearly just society has full authority, which means that the citizens of such societies have content-independent duties to obey the law. Despite having full authority, states in nearly just societies can

nevertheless enact unjust laws. The citizens of such societies have duties/reason to encourage their states to change these laws, and such encouragement may sometimes need to take the form of radical activism. This creates the possibility of a conflict between full authority and the duty to obey the law, on the one hand, and the duty/reason to remedy social injustice, on the other, with full authority effectively constraining the practice of radical activism. Similarly, in conditions of weak authority, citizens face the possibility of a conflict between the duty to avoid acts that risk re-entry into the state of nature and the duty/reason to remedy social injustice, with weak authority effectively constraining the practice of highly disruptive radical activism.

I want to suggest that citizens of seriously unjust regimes who are contemplating engaging in highly disruptive radical activism must practise restraint if they are targeting a regime that has weak authority, but need not practise restraint if they are targeting a regime totally lacking in authority. This restraint manifests itself in a commitment to avoid highly disruptive activism. For an illustration, imagine there are two societies, both severely unjust and so lacking full authority. In one society, the state claims the right to rule but has in fact failed to secure order and security in the territory it claims authority over.[43] Call this society *Disorder*; its regime lacks all authority. In the other society – call it *Order* – the state has secured order and security and so has weak authority. Now imagine that there is a form of highly disruptive radical activism – a short insurgency, for instance – that would remedy injustice in both societies, and that the citizens of both societies contemplating such activism are equally confident that the success, necessity, and proportionality conditions have been met.

My claim is that citizens of *Order* should initially practise restraint – avoiding highly disruptive activism they are confident is justified – while the citizens of *Disorder* need not. This does not mean that citizens of *Order* are never permitted to resort to highly disruptive activism. Rather, it means that they must proceed with more caution than if they were living under a regime totally lacking in authority. Effectively, at least when it comes to highly disruptive radical activism that seeks to frustrate and undermine the ability of the state to maintain order, the threshold for 'sufficient confidence' is higher when targeting regimes that have weak authority compared to those totally lacking in

[43] Potential real-world examples of such regimes include the Libyan Government of National Unity, the Cabinet of Yemen, or the Government of the Syrian Arab Republic.

authority. This means that the constraints on the practice of radical activism imposed by weak authority operate via the conditions for permissible defensive harm: the epistemic criteria required to satisfy the success, necessity, and proportionality conditions are harder to meet under weak authority than they are under no authority.

While I do not want to suggest that this view in any way informed their strategy, it seems to me that the account of the relationship between authority and radical activism that I am developing can uncover and bring to light a plausible normative rationale for the approach that the African National Congress (ANC) and its armed wing, Umkhonto we Sizwe (Umkhonto), took to resisting apartheid in South Africa, at least according to the account Nelson Mandela provides in the statement he gave in his defence during his 1964 trial (Mandela 1964). Apartheid South Africa was a severely unjust regime of legally enforced white supremacy; none of its citizens had a duty to obey the law as such. Nevertheless, it is at least arguable (though certainly not uncontroversial) that this regime managed to maintain order, for the most part through threats and force alone. Using my terminology, it lacked full authority but had weak authority.[44]

Citizens of apartheid South Africa initially attempted to remedy the unjust circumstances they faced via ordinary activism, but these efforts failed,[45] in large part because the black majority were prohibited from engaging in many forms ordinary

[44] Mandela seems to endorse a binary view of political authority and ground the right to rule in the will of the people: "The African people were not part of the Government and did not make the laws by which they were governed. We believed in the words of the Universal Deceleration of Human Rights, that 'the will of the people shall be the basis of the authority of the Government'" (Mandela 1964). I mentioned earlier (see note 38) that the taxonomy of political authority I provide in this chapter is not intended to be exhaustive, and nothing I say here should be taken to deny that the will of the people is relevant to the authority of a regime. The whole point of disaggregating political authority is to show that there is more than one authority relation that might exist between a state and its citizens (and between either and outside agents), and that these relations can have different grounds.

[45] Compare Mandela: The African National Congress was formed in 1912 to defend the rights of African people which had been seriously curtailed… For thirty-seven years – that is until 1949 – it adhered to a strictly constitutional struggle. It put forward demands and resolutions; it sent delegations to the Government in the belief that African grievances could be settled through peaceful discussion and that Africans could advance gradually to full political rights. But White Governments remained unmoved, and the rights of Africans became less instead of becoming greater" (Mandela 1964).

activism.[46] At this point, given the egregiousness of the injustice and the intransigence and past behaviour of the regime, I do not think it would have been unreasonable for a citizen surveying the scene to conclude that highly disruptive radical activism such as civil war was permissible, and many citizens did in fact come to such a conclusion.

Nevertheless, the ANC sought to practise restraint.[47] This restraint took the form of first engaging in illegal but nonviolent activism[48] and, when that failed, then engaging in forms of violence,[49] such as sabotage, that do not seek to undermine the ability of the regime to maintain order.[50] One of the key motivations for showing restraint was the desire to avoid civil war.[51] Nevertheless, though civil war was to be avoided, it was not to be ruled out.[52] It seems, then, that the ANC/Umkhonto took their society to be

[46] Compare Mandela: "All lawful modes of expressing opposition to [white supremacy] had been closed by legislation" (Mandela 1964).

[47] Compare Mandela: "for a long time the people had been talking of violence… and we, the leaders of the ANC, had nevertheless always prevailed upon them to avoid violence and to pursue peaceful methods" (Mandela 1964).

[48] Compare Mandela: After 1949 "there was a change from the strictly constitutional means of protest which had been employed in the past. The change was embodied in a decision which was taken to protest against apartheid legislation by peaceful, but unlawful, demonstrations against certain laws" (Mandela 1964).

[49] Compare Mandela: "It was only when all else had failed, when all channels of peaceful protest had been barred to us, that the decision was made to embark on violent forms of political struggle, and to form Umkhonto we Sizwe" (Mandela 1964).

[50] Compare Mandela: "Four forms of violence were possible. There is sabotage, there is guerrilla warfare, there is terrorism, and there is open revolution. We chose to adopt the first method and to exhaust it before taking any other decision… Sabotage did not involve loss of life, and it offered the best hope for future race relations. Bitterness would be kept at a minimum and, if the policy bore fruit, democratic government could become a reality" (Mandela 1964).

[51] Compare Mandela: "We did not want interracial war, and tried to avoid it to the last minute" (Mandela 1964).

Also: "We felt that the country was drifting towards civil war in which Blacks and Whites would fight each other. We viewed this situation with alarm… with civil war, racial peace would be more difficult than ever to achieve" (Mandela 1964).

[52] Compare Mandela: "The avoidance of civil war had dominated our thinking for many years, but when we decided to adopt violence as a part of our policy, we realized that we might one day have to face the prospect of such a war. This had to be taken into account in formulating our plans… We did not want to be committed to civil war, but we wanted to be ready if it became inevitable" (Mandela 1964).

characterised by order – the obverse of civil war – and practised restraint at least in part because, while they took civil war to be in principle permissible, they wanted to avoid the seriously undesirable implications of disorder. Using the terminology I propose here, the practice of radical activism was constrained by the presence weak authority.

This account of the ethics of radical activism is more restrictive than the binary, justice-grounded view. According to the latter, radical activism in seriously unjust societies is not at all controversial at the bar of authority. The justifiability of radical activism is determined by appeal to the conditions for permissible defensive harm and authority plays no part in these considerations. My view, however, is that regimes that are seriously unjust can nevertheless possess weak authority and that this authority status imposes constraints on the practice of radical activism, constraints which operate via the conditions for permissible defensive harm.

3.5. Conclusion

In this chapter, I have argued that the binary, justice-grounded account of political authority that is often presupposed in discussions of the ethics of radical activism should be disaggregated, and one upshot of the disaggregation I propose is a conception of political authority I called weak authority. I also argued that one implication of recognising the existence of weak authority is that the practice of radical activism is more morally constrained than is typically assumed. Having focused on law as the site of social justice activism in this chapter, in the next chapter I explore the question of how to intervene in social norms.

4. Social norms, civil society, and the ethics of informal activism

> Some injustices are remediable only through action in civil society.
>
> Anderson 2009:144

Introduction

In the previous chapter, I explored the ethics of activism that seeks to remedy the form of social injustice with which we are most familiar: injustice underpinned by law. Given that laws are formal rules (see chapter 2), I called activism that seeks to change the law formal activism. However, we also saw in chapter 2 that injustice can be generated by social phenomena other than law. One such phenomenon is a social norm. Recall that a social norm is a rule that people prefer to comply with on the condition that they believe that: (a) most members of their reference network comply with the rule, and (b) most members of their reference network have personal normative beliefs that correspond to the rule.[53] To have a personal normative belief that corresponds to a rule is to believe that people have an obligation to comply with the rule and to be disposed to punish those who do not comply with it (Bicchieri 2017:35).

Plausible examples of social norms which generate social injustice include: the rule which exists in some African communities which prescribes female genital cutting (FGC); the rules which exist in some Indian communities which legitimise caste discrimination; the rules which exist in some Arab and Muslim communities which underpin the practice of 'honour killing'; the rule which exists in many western countries according to which the sexual harassment of women should be tolerated; and the rule which exists in communities the world over which requires women to carry out the majority of unpaid domestic labour. In each of these cases, some – perhaps many – of those who comply with the relevant rule may do so because they believe that most other

[53] It is worth reiterating that this definition of a social norm is a "rational reconstruction" (Bicchieri 2006:3, 10), which means that it should not be read as a faithful description of the beliefs and desires people consciously have but as a way of uncovering the basic elements of social norms which typically lie outside of conscious awareness. The key desiderata of a rational reconstruction are explanatory power and testability.

members of their reference network comply with the rule, and that most other members of their reference network have personal normative beliefs that correspond to the rule.

The observation that social norms can generate social injustice combined with the view (defended in chapter 1) that citizens of unjust societies have duties to remedy the injustice in their midst yields the conclusion that citizens have duties to remedy injustice generated by social norms. Remedying injustice generated by social norms will involve encouraging people to abandon the prevailing unjust social norm and perhaps also to comply with a new, more just one instead. As I suggested in chapter 2, because social norms are informal rules, let us call this type of activism informal activism. This chapter explores how informal activism should be practised.

I have two aims in particular. The first – pursued in sections 4.1 and 4.2 – is to critique a widely assumed approach to social justice activism which may be thought sufficient to answer the question of how informal activism should be practised. According to this approach, the state is the primary agent of social justice and so citizens' duties to remedy social injustice must be discharged by encouraging the state to change law and policy. I argue that this approach relies on a simplistic understanding of the relationship between laws and social norms. Moreover, it fails to appreciate the fact that the efficacy of the state's efforts to remedy social injustice caused by social norms will be constrained by the fact that there is a realm of social life within which legitimate state action is limited. I will refer to this realm as 'civil society'. Given the limits imposed on state action by civil society, it will often be the case that citizens will need to discharge their duties to remedy social injustice via means that do not involve encouraging the state to change law and policy.

Having highlighted the limits of state action, my second aim – pursued in sections 4.3-4.6 – is to explore how informal activism in civil society should be practised. I argue that the value that should guide the practice of informal activism in civil society is the very same value that informs civil society itself: respect for moral personhood. This value generates three key considerations pertaining to informal activism in civil society. The first is that persuasion is the ideal means of influence in civil society. This means that the preferred way for informal activists to remedy social injustice and advance social justice is for them to offer others good reasons to change their beliefs or behaviour. The second consideration is that there is a presumption against what I call 'pressure'. This means that it is presumptively impermissible for informal activists to seek to change others'

behaviour by physically forcing them to do so or by making it so that their fundamental interests are at risk if they don't. The third consideration is that what I call 'price-setting' is nonideal but nevertheless permissible. By 'price-setting' I mean efforts to change others' behaviour through the use of incentives where fundamental interests are not at stake. I argue that informal activists may engage in price-setting, but there are both principled and practical reasons which speak against this as the preferred form for informal activism to take. Section 4.7 concludes.

4.1. Social justice, the state-mediation thesis, and the unjust social norm objection

It might seem that the answer to the question 'how should informal activism be practised?' is quite simple. It is tempting to think that informal activism should be practised in the same way that all social justice activism should be practised: by encouraging the state to change law and policy. My aim in this section is to explore the rationale for this claim, to introduce a prominent objection to it, and to highlight how defenders of the original claim have responded to this objection.

4.1.1. The state-mediation thesis

According to a widely held view, the state is the *primary* agent bearing *direct* duties to realise social justice and ordinary citizens must, therefore, play an *indirect, secondary* role (see O'Neill 2001). This means that, when it comes to promoting justice and remedying injustice, ordinary citizens should not seek to do so directly via actions that are not mediated by the state. Instead, ordinary citizens' duties to promote justice and remedy injustice must be discharged by encouraging the state to reform and/or enforce law and policy. Let us call such activism *state-mediated activism* and let us call the view that ordinary citizens' duties to promote justice and remedy injustice should be discharged via state-mediated activism *the state-mediation thesis*.

Many take John Rawls to be endorsing the state-mediation thesis when he says that principles of justice apply directly only to a political community's major social institutions taken together as one scheme – which he calls the "basic structure of society" – and not to the actions of individual citizens. Instead of being required to realise justice directly, individuals have a "natural duty" of justice "to support and to comply with just institutions that exist and apply to us" and "to further just arrangements not yet established" (Rawls 1999a:99). Rawls's use of the phrase 'basic structure of society' is

often taken to refer to the legal system of the coercive state (e.g., James 2005, Meckled-Garcia 2008). On this reading, Rawls's view is that individuals have duties to promote justice that should be discharged, not in their unmediated interactions with each other, but *via state institutions*. This means either supporting the state and complying with the law when society is just – or "nearly just" (Rawls 1999a:319) – or demanding that state officials change law and policy when society is unjust.[54]

A plausible rationale for the state-mediation thesis is capacity-based. Given that social justice refers to society-wide patterns of relationships and entitlements, the primary responsibility for realising social justice must fall on an agent able to monitor, regulate, and adjust society-wide relationships and entitlements so as to secure and maintain whichever pattern we define as just. To illustrate, consider a relatively *laissez faire* conception of social justice according to which individuals' entitlements and relationships should be determined by whichever agreements they come to with each other, so long as these agreements are free and fair. Over time, the cumulative effects of many free and fair exchanges are likely to result in a distribution of resources and opportunities such that the conditions for free and fair agreements no longer obtain. This means that it is necessary to monitor, regulate, and continually adjust the distribution of resources and opportunities to ensure that the conditions for free and fair agreements persist.

But even well-meaning individuals could not hope to successfully design their actions so as to guarantee any particular society-wide distribution of resources and opportunities (Rawls 2005:265-9). Only an agent able to direct and coordinate the actions of all other agents in society is able to perform such a task. The paradigmatic

[54] Despite the fact that Rawls is often regarded as an advocate of the view that social justice activism should be state-mediated, it actually seems to me that his position is more ambiguous. The first half of the natural duty of justice specifies what citizens should do when their society is just or nearly just, and let us assume, as is common, that by 'institutions' here Rawls means to refer to the legal system of the coercive state. The second half of the natural duty of justice refers to what citizens should do when their society is unjust. It is noteworthy, however, that Rawls replaces the word 'institutions' with the word 'arrangements' when referring to what it is that citizens should 'further' in conditions of injustice. This hints at the possibility that Rawls does not assume that social justice activism should be state-mediated, as he is typically taken to, but actually has a more capacious view of the forms that social justice activism might take. I will not explore this possibility any further here but will instead proceed on the basis that Rawls should be read as an advocate of state-mediated activism.

example of such an agent in the modern world is the state and the principal means by which the state directs and coordinates the actions of its citizens is by issuing commands in the form of laws. This line of reasoning in defence of the state-mediation thesis applies *a fortiori* for conceptions of social justice that are not *laissez faire* but avowedly interventionist.

This capacity-based rationale is further strengthened by the fact that, as I mentioned in the previous chapter, social injustice is often generated by the laws and policies of the state. When this is the case, the unjust laws and policies in question will need to change if social injustice is to be remedied. Recall, however, that law is best characterised as a system of rules which includes secondary rules that empower state officials to, amongst other things, change the rules of the system (Hart 2012). This means that, when it comes to remedying social injustice generated by law, ordinary citizens (i.e., non-officials) should not discharge their duties to engage in social justice activism *directly* by seeking to change the law themselves, for it is not within their power to do so. Instead, ordinary citizens should seek to remedy social injustice generated by law *indirectly*, by encouraging state officials to change the law.

4.1.2. The unjust social norms objection

Despite this plausible rationale, the fact that social norms can generate social injustice poses a challenge for the state-mediation thesis. This challenge is most famously presented by G. A. Cohen (1997). Cohen objects to Rawls's claim that principles of justice apply directly only to 'the basic structure of society' and that ordinary citizens' justice-related duties must be discharged via this structure. Assuming that the basic structure refers to the legal system of the coercive state, Cohen notes that social phenomena that reside 'outside' of the basic structure can generate social injustice. He cites the example of social norms which determine how power, opportunity, and labour are distributed between men and women in heterosexual households. The fact that some ordinary citizens comply with social norms which require women to shoulder a much greater share of the domestic responsibilities compared to men, or demand that sons are encouraged to pursue higher education while daughters are not, is surely something that a theory of social justice should be concerned with.

But, Cohen observes, the claim that principles of justice apply directly only to 'the basic structure' – that is, the state and its legal system – and that ordinary citizens' only

justice-related duties are state-mediated, appears to yield the implausible implication that social norms and the behaviour they generate are irrelevant to the pursuit of social justice. Against the state-mediation thesis, Cohen argues that state action and state-mediated activism are not sufficient for the realisation of social justice; ordinary citizens must also promote justice via their unmediated interactions with each other.[55] Call this *the unjust social norms objection* to the state-mediation thesis.

Some have sought to defend the state-mediation thesis from the unjust social norms objection by insisting that justice requires the state to enact laws and policies that are sensitive and responsive to the prevailing social norms of a society and that seek to counteract or change those social norms if they are thwarting the realisation of social justice (Ronzoni 2008, also see Ronzoni 2007).[56] It follows from this that social norms are far from irrelevant to the pursuit of social justice. Let us call efforts by the state to counteract or change unjust social norms *state-led social norm management* (see Sunstein 1996). State-led social norm management can take a variety of forms, including: education and advocacy campaigns; using economic instruments to tax or subsidise compliance with social norms; imposing time, place, and manner restrictions on social norm-governed behaviour; and legally prohibiting (or requiring) the behaviour required (or prohibited) by social norms (Sunstein 1996:948-952).

4.2. Civil society and the limits of state-led social norm management

I believe that law and policy have a very important role to play when it comes to counteracting or changing unjust social norms. Nevertheless, I also believe that state-led social norm management is unlikely to be sufficient to realise social justice, and so

[55] Cohen does not frame his argument in terms of what I call the state-mediation thesis. Instead, he takes himself to be arguing, against Rawls, that principles of justice apply "to the choices that people make *within* the legal coercive structures to which, so everyone would agree, principles of justice (also) apply" (Cohen 1997:3). However, given that Rawls argues that principles of justice do apply to individuals' choices in the form of the natural duty of justice (and Cohen acknowledges this), I take it that the dispute between Cohen and Rawls boils down to whether ordinary citizens' duties to promote justice and remedy injustice should be discharged via the 'legal coercive structure' – i.e., the state – or not.

[56] Like Cohen, Ronzoni also does not frame her argument in terms of the state-mediation thesis. However, the view that she defends against Cohen's objection is equivalent to the state-mediation thesis. This view is that primary, direct duties to realise justice apply only to the basic structure of society and that ordinary citizens' only justice-related duties must be discharged via this structure.

ordinary citizens' efforts to promote justice and remedy injustice should sometimes not be meditated by the state. This, recall, is the essence of the unjust social norms objection to the state-mediation thesis. My aim in this section is to highlight the limits of state-led social norm management, and, in so doing, to vindicate the unjust social norms objection.

4.2.1. The importance of civil society

A key reason why the efficacy of state-led social norm management is limited is that much of the social norm-governed behaviour of ordinary citizens that is relevant to the realisation of social justice takes place in an area of social life where the state's ability to legitimately intervene is restricted. Without providing a full theory or defence of this protected area of social life, let me give some substance and plausibility to the idea by briefly sketching the outlines of an account.

The first thing to do is to give this protected area a name; I will call it *civil society*. I am aware that this term has been and is used to denote different things in different contexts, but given my current aims this labelling seems to me to be appropriate.[57] Nevertheless, there is a sense in which my use of this term is somewhat stipulative. 'Civil society' just is the label that I will use to refer to the protected area of social life which I think limits the state's ability to engage in social norm management.

It seems to me that recognition of the importance of civil society is one of the implications of taking seriously a certain conception of human beings. According to this conception, human beings are creatures who are not simply able to reason instrumentally, and so set and pursue plans to realise their wants, needs, and desires in light of prudential considerations. Instead, in the overwhelming majority of cases, fully-developed, adult human beings also have the capacity to understand and respond to *moral reasons*, to set (and reset) long-term goals and to organise (and reorganise) their life plans in light of these moral reasons, and to be motivated to pursue these long-term goals and life plans even when they do not align with their immediate wants, needs, and desires (Gaus 2011:202-5). Call this latter feature of human beings our *moral personhood*. I assume that it is in virtue of the fact that human beings are moral persons that they are

[57] The sketch of civil society I offer here aligns most naturally with the perspective on society/state relations which Simone Chambers and Jeffrey Kopstein call "civil society apart from the state" (Chambers and Kopstein 2006:364-367).

owed a special kind of respect that is not owed to (most) other animals. Moreover, I also assume that being a moral person is a sufficient condition for being subject to principles of social justice and for having basic rights (Rawls 1999a:441-9).

While the conception of human beings just highlighted undergirds the account I favour, civil society is given its distinctive shape and structure by the fact that moral persons have certain fundamental interests which give rise to certain basic rights and duties. Let me highlight three in particular. The first is the interest in having and honouring *conscientious moral convictions*. Conscientious moral convictions are important because they are the serious and sincere moral judgements that are the end products of the exercise of moral reasoning. To have and honour such convictions is to live one's life in accordance with them: to comply with them and be motivated to comply with them, to believe that others ought to comply with them in the relevant circumstances, and to be willing to communicate them in an effort to engage others in reasoned dialogue about their merits. Kimberley Brownlee argues persuasively that our interest in having and honouring our conscientious moral convictions should be protected by rights to freedom of thought and freedom of expression (Brownlee 2012:128-141).

The second fundamental interest is in having *intimate social connections*. These are relationships which persist over time, are constituted by recurring interactions with particular individuals, and which we tend to identify with and be deeply invested in. Intimate social connections are important for a number of reasons. For one thing, they are intrinsically valuable; they are partly constitutive of a minimally decent human life (Griffin 1988:68-9). But they are also instrumentally valuable for they are necessary for the enjoyment or pursuit of other goods. Examples include the resilience to withstand setbacks and hardship, a subjective sense of belonging, and a social network able to enforce rules of appropriate behaviour (Wolff and de-Shalit 2007:138-142).

One of the most important goods that intimate social connections facilitate access to is the having and honouring of conscientious moral convictions. Intimate social connections are often the vehicles through which we comply with and communicate our conscientious moral convictions; religious communities are perhaps the paradigmatic example of such a phenomenon. But intimate social connections are also a necessary precondition for having and honouring conscientious moral convictions. Without the persistent contact and care typically provided by families, for example, children are

unlikely to develop the cognitive and emotional capacities that are prerequisites for moral reasoning (Brownlee 2013:210, 2015:275, 2016:64).

Therefore, not only does our interest in intimate social connections ground a right to freedom of association – subject to certain constraints (see Brownlee 2015) – but it also grounds moral duties to maintain intimate associations with those who are dependent on us. The most obvious example of such a case are the duties that parents have to their young children, but other examples include the duties that adults have to elderly and/or severely impaired friends or relatives (Brownlee 2015:272, 275-6). In fact, citing the example of "the associations among the 33 men in the Copiapó mining accident in Chile that trapped them underground for 69 days", Brownlee argues that we may even have duties to *form* – not merely maintain – intimate associations if the circumstances made such intimate association necessary (Brownlee 2015:273).

The third fundamental interest underpinning civil society is *control over how we present ourselves to others*. This interest is important for three reasons. The first is that being able to control how we present ourselves to others enables us to create and maintain different kinds of relationships, in particular the intimate associations just highlighted. This is because different patterns of social behaviour are associated with different relationships; being a spouse or a friend or a teacher or a colleague each requires that we behave in different ways and share different information. As James Rachels notes, "our ability to control who has access to us, and who knows what about us, allows us to maintain the variety of relationships with other people that we want to have" (Rachels 1995:329). Second, and relatedly, because intimate knowledge about others often comes with expectations of care, a certain level of social detachment enables us to manage, at least in part, the responsibilities we undertake. The third reason why having control over how we present ourselves to others is important is that we may not want to expose aspects of ourselves to social scrutiny, especially if that social scrutiny is likely to come with social costs. Invoking these reasons, Andrei Marmor argues that our interest in having control over how we present ourselves to others grounds a right to privacy (Marmor 2015).[58]

[58] See Nagel (1998) for an insightful and illuminating discussion of the nature, role, and dynamics of privacy in individual and social life.

Together, these three fundamental interests and the rights and duties they generate establish the shape and dynamics of a protected area of social life – civil society – in which ordinary citizens ought to be able to control how they present themselves to others, to form and maintain intimate associations, and to develop and communicate their deeply held moral commitments. Indeed, these three activities are often intertwined. Ordinary citizens will exercise control over how they present themselves to others *in order to* form and maintain intimate associations *with those who* share their deep moral commitments, associations *within which* these deep moral commitments are honoured, instantiated, propagated, and re-affirmed. Examples of practices in which these three activities are related in this way include friendships and religious associations, but perhaps the paradigmatic case is the family. Given this, let me illustrate the limits of state-led social norm management by focusing on the example Cohen himself cites: sexist social norms relating to the distribution of power, opportunity, and labour in heterosexual households.

If enough people who endorse these social norms choose to form families in which they comply with them and socialise their children to accept them, then these social norms will continue to circulate in society, and there will be a limit to how much the state can do to realise social justice in such circumstances. Of course, the state might engage in various forms of social norm management. Interventions that have been suggested in response to Cohen's objection include creating incentives for women to work outside the home and for men to care for children, and using the public education system to inculcate egalitarian values in children (Schouten 2013:380-1). Other proposals might be for the state to provide girls with ample – perhaps additional – opportunities via the public education system, and to ensure that workplaces are compatible with the domestic responsibilities that women are expected to discharge at home (Anderson 2009:141). However, so long as citizens enjoy the protections that civil society provides, the state's ability to effectively counteract sexist socialisation or undermine sexist social norms will be seriously limited. "Some gender injustice is thus liable to remain in a liberal society, even when everyone acts within their legal rights and the state does all it can legitimately do to promote gender justice" (Anderson 2009:141).

There are both issues of feasibility and issues of desirability at play here and so it is worth teasing them out and being explicit about how they relate to each other. One might argue that, beyond the measures suggested in the previous paragraph, state-led

social norm management is simply unlikely to be effective. For instance, referring to state action to limit the expression and propagation of sexist or patriarchal social norms, Elizabeth Anderson says that this "would entail immense costs, while still being unlikely to change people's minds" (Anderson 2009:143). I want to make a different point. My argument is not that, beyond the kinds of measures discussed in the previous paragraph, no intervention by the state would effectively undermine unjust social norms. Examples of interventions that seek to limit the expression and propagation of sexist and patriarchal social norms, which go beyond the suggestions made in the previous paragraph, include: prohibiting the expression or promotion of such norms within families; setting up systems to monitor and surveil individuals' behaviour within families to ensure that they do not express or promote such norms or they are effectively punished when they do; or removing children from families to ensure that they are raised in an environment in which they are not socialised to endorse sexist and patriarchal social norms. I am happy to grant for the sake of argument that the state may sometimes be able to intervene in these ways and that such measures may sometimes successfully undermine unjust social norms. In this sense, these interventions can be feasible.

However, my claim is that, despite sometimes being feasible, these kinds of interventions are always undesirable and in fact impermissible – and they are impermissible because individuals have the rights which create civil society identified earlier. Hence, my view can be presented as a feasibility claim that is conditional on a desirability claim: *if* the protected space that is civil society is recognised and respected as it should be, *then* the state's capacity to successfully undermine unjust social norms is severely limited.

This doesn't mean that injustice generated by social norms must be tolerated or that that there is nothing that can be done about it. Instead it means that the rationale for the state-mediation thesis – that citizens should discharge their justice-related duties via the state because the state has a special capacity to realise social justice – is undermined. Sometimes ordinary citizens will need to promote justice and remedy injustice, not just via state-mediated activism, but also through actions that take place in civil society, in ways that are not mediated by the state. And this, recall, was the conclusion of the unjust social norms objection to the state-mediation thesis.[59]

[59] Compare Swift:

4.2.2. What about social norms that threaten fundamental interests?

One possible rejoinder to what I have said so far would be to concede that the efficacy of state-led social norm management is limited when it comes to the kinds of social norms Cohen highlights, but to deny that this is the case for unjust social norms generally. What is distinctive about social norms according to which women should undertake more domestic labour than men, or sons should be encouraged to pursue higher education while daughters should not, is that they prescribe behaviour that is unjust but nevertheless should presumably be legally permissible. However, many unjust social norms require or legitimise behaviour that attacks or threatens people's fundamental interests and so which any theory of social justice would demand that law prohibits. This is true of the social norms underpinning FGC, 'honour killing', or caste discrimination, for instance. These activities should be criminalised independently of whether social norms support them and so, in these cases, state-led social norm management should take the form of legal prohibition. One might even think that, by criminalising the behaviour in question, the state would be issuing an authoritative repudiation of the unjust social norm which itself might encourage citizens to revaluate and eventually abandon the social norms underpinning that behaviour.

This is a very important observation which I would not want to deny. There are, however, four points it is worth recording which qualify this observation somewhat. First, the rejoinder presupposes that the (criminal) law is always an effective means by which to regulate social behaviour, but this is not necessarily the case. This is because effective rule of law itself requires a social norm of legal obedience, which is not always present (Mackie 2017). The post-Soviet regimes of Eastern Europe provide a striking example. According to Denis J. Galligan, these are societies in which "citizens, organizations, businesses, and even government officials organize their affairs with scant regard for the law" (Galligan 2003:1). This state of affairs does not necessarily reflect the personal normative beliefs of post-Soviet citizens; they are just as likely to endorse the rule of law as citizens of western regimes are (Galligan 2003:1, Kurkchiyan 2003:31).

> Even if one thinks that people should devote most of their energy to changing policies so that they better promote justice, and that justice requires little of people as individuals other than to play their role in bringing about political change or change at the policy level, it is unlikely that it demands *nothing* else of them (Swift 2008:379 italics in original).

Instead, the widespread disregard for the law is a legacy of Soviet (and pre-Soviet) rule in which law was seen as an alien instrument of coercion, and so reciprocal expectations of getting around the law emerged. These reciprocal expectations have persisted despite the change in political regime (Kurkchiyan 2003).

The phenomenon of disregard for the law is not limited to post-Soviet societies; it is a problem for some postcolonial regimes too (Mackie 2017:319). And it may help to explain why, though most African states have legally prohibited FGC, the practice is still widespread, but rates are lower when trust in government institutions is high (Bicchieri and Marini 2015:21). Hence, it seems that, when there is no social norm of legal obedience, criminalisation alone is unlikely to be an effective means by which to advance social justice. Social norms themselves will need to change and this change cannot be secured via law.

The second point is that, even where there is a social norm of obedience to law in general, criminalisation is likely to fail when a particular law conflicts sharply with prevailing social norms. This is because criminal law needs to be enforced, and those charged with enforcing the law – police officers, prosecutors, juries, and judges – are reluctant to enforce laws which depart too far from social norms which they endorse (Kahan 2000). This reluctance to enforce may even have the perverse effect of reinforcing the very social norms the law seeks to condemn, by signalling to the public that the practices in question are widely accepted.

This phenomenon may help to explain why caste discrimination in India is still prevalent despite the fact that it is declared illegal by the Indian Constitution. According to Jeremy Sarkin and Mark Koenig, "anti-discrimination efforts have been impotent through poor enforcement and general lack of political will. The police and low level judiciary have been at the center of the failure to prevent caste discrimination" (Sarkin and Koenig 2010:550 reference removed). The phenomenon of laws being in place but not enforced may also explain why the practice of 'honour killing' persists in Pakistan (see Appiah 2010:137-172). Hence, in scenarios in which unjust social norms are widely endorsed by those charged with enforcement, simply criminalising the undesirable behaviour is unlikely to be effective. Instead the social norms themselves will need to change.

There is an important caveat to this point. According to Dan M. Kahan, law can sometimes be effective at changing social norms, but only if it proceeds by way of "gentle

nudges" which depart slightly from the prevailing social norms, not "hard shoves" which entail severe condemnation. First, a mildly condemnatory law is enacted and enforced, which leads ordinary citizens to revise their attitudes towards the behaviour in question in the direction of the law. This opens up space for a further, slightly more condemnatory law to be enacted, which leads citizens to revise their attitudes once again. And so on. Eventually, we would arrive through incremental steps at a situation in which the unjust social norm has been totally eroded. But, crucially, this result would not have been achieved had we simply begun by implementing the ultimately desired legislation without first taking the incremental steps. Kahan argues that this model helps to explain the history of the evolving relationship between laws and social norms pertaining to date rape, smoking, domestic violence, drugs, drink driving, and sexual harassment in the USA (Kahan 2000).

However, even if we were to take Kahan's advice and proceed by way of gentle nudges, not hard shoves, when it comes to reforming an unjust social norm, there would still be an extended period during which the norm in question – or slightly less egregious versions of it – prevails. What should citizens do in the meantime? In this case, the state-mediation thesis would have us think that citizens' efforts to undermine the unjust social norm should be focused exclusively on securing incremental changes to law and policy. But this seems to make a fetish of legislative change, especially now that we see that legislative change is not a straightforward panacea.

Moreover – and this is the third point to make in response to the rejoinder we are considering – in regimes that are at least minimally responsive to public opinion,[60] an important means by which to secure changes in legislation is by changing social norms. Consider the American civil rights movement of the 1960s. The principal audience for the campaign for the Voting Right Act and Civil Rights Act was not the federal government, but white citizens of the northern states. The aim was to change their personal normative beliefs regarding what was acceptable and in so doing put pressure on the federal government to act. Recall that social norms are rules that are complied with, in part, because those who comply with them believe other members of their reference network have personal normative beliefs that correspond to the rule. Changing social norms was the means by which legislative change was secured.

[60] I take this to include well-functioning democracies, but not only them.

The fourth and final point is that, though limited, the concession embedded within the rejoinder we are considering is itself highly significant. Even if it was the case that the efficacy of state-led social norm management was 'only' limited with respect to social norms supporting behaviour that is unjust but should nevertheless be legally permissible – and I have tried to suggest that this is not the case – this would still seriously undermine the state-mediation thesis. This is because social norms prescribing unjust but legally permissible behaviour are an important source of social injustice in many modern societies. Sexist social norms pertaining to the role of women in the home and wider society provide just one set of examples. Homophobic social norms according to which homosexuality should be regarded as a source of shame provide another.

Together, the four points recorded in this subsection suggest that, when it comes to remedying injustice generated by social norms, there is a vitally important role for informal activism in civil society to play. My aim in the following four sections is to explore how informal activism in civil society should be practised.

4.3. The importance of critical movements

Informal activism in civil society will typically begin with the formation of what I will call *critical movements*.[61] Critical movements are groups of individuals who seek to challenge prevailing social practices, and to develop and disseminate new practices to replace existing ones. Unsurprisingly, these groups will normally be made up of individuals who cannot see many good reasons to adhere to the widespread patterns of behaviour that they wish to challenge and replace – that is, people who are less sensitive to these practices. However, critical movements also tend to be populated by people who have greater levels of autonomy and perceived self-efficacy, and who have low sensitivity to

[61] Thomas R. Rochon distinguishes between what he calls "critical communities", which are groups of thinkers who question, challenge, and propose alternatives to the status quo, and "movements", which are networks of individuals who take up and disseminate the ideas and arguments of critical communities, reframing and reformulating these ideas so that they are more likely to win widespread acceptance (Rochon 1998:22-53). Rochon argues that while this distinction between critical communities who generate ideas and movements who disseminate them is useful for analytical purposes, the two groups and processes are often empirically intertwined. In sketching critical movements, I have drawn on and blended Rochon's descriptions of both critical communities and movements.

and perception of risk.[62] Critical movements can vary in size but, given that they are self-consciously transgressive, they will initially be small relative to the community of people who currently manifest the patterns of behaviour the members of the critical movement wish to see changed.

As I argued in chapter 2, there are at least three kinds of social phenomena that might underpin unjust patterns of behaviour – stereotypes, social norms, and law – and critical movements might be formed to target any of these mechanisms. Indeed, some critical movements seek to intervene in multiple mechanisms. Nevertheless, given the topic of this chapter, I will focus exclusively on those critical movements or those aspects of critical movements which target social norms specifically. Let us call the members of critical movements which target social norms *informal activists*, and non-members of such critical movements – those who the members of critical movements seek to influence – *nonactivists*. Historical examples of critical movements include the British abolitionists of the late 18th and early 19th centuries and the women's liberation movement of the 1960s and 70s.

Critical movements are important for informal activism and informal activists for a number of reasons. Let me highlight three. First, critical movements serve what I will call a *protective function*. Recall that noncompliance with social norms is likely to be met by sanctions. Because they tend to be more autonomous and have lower levels of risk sensitivity and perception, informal activists are likely to be the members of the community most able to resist the pressure from others to conform. They will, nevertheless, want to avoid or mitigate the worst negative consequences of deviance, and forming critical movements can help them to do so. This is the sense in which critical movements are 'protective'.

Consider, for instance, the case of footbinding in China. Footbinding was the often extremely painful practice of wrapping the feet of young girls in cloth to modify their shape and size, making them smaller than they would naturally be. It began around the late 10th century as a marker of high status and it spread throughout Chinese society over the centuries, though it remained concentrated in the upper classes. Despite the pain and

[62] Focusing on social norm-governed practices in particular, Cristina Bicchieri argues that possession of these features – low sensitivity to the norm, high levels of autonomy and perceived self-efficacy, and low sensitivity and perception of risk – are the makings of what she calls a "trendsetter" – that is, an individual who is more likely to move first to abandon a prevailing social norm (Bicchieri 2017:165-182).

damage the practice entailed, bound feet were widely considered beautiful and desirable. As Kwame Antony Appiah reports, "[w]omen with natural feet were mocked; women with small feet… were praised and prized" (Appiah 2010:70). In particular, having bound feet came to be regarded as a condition of marriageability. Women with natural, unbound feet would face difficulties in finding a partner and so families had a strong incentive to ensure their daughters' feet were bound. In order to protect themselves against some of the negative consequences of abandoning the practice, those who rejected footbinding formed 'natural foot societies', the members of which pledging to both not bind their daughters' feet and not allow their sons to marry women with bound feet (Mackie 1996, Appiah 2010:85-7).

Second, critical movements are *sites of innovation*. By this, I mean to refer to the role that critical movements play as fora within which the ideas, arguments, and tactics for challenging prevailing social norms can be shared, scrutinised, and developed. Consider, for example, the phrase 'sexual harassment', which entered into the American vocabulary in the mid-1970s. Through intra-movement discussion and debate via conferences, articles, and books, this phrase was developed within the feminist community to capture and crystalize their critique of prevailing workplace practices in the United States (Rochon 1998:68-79). While the phrase 'sexual harassment' was first deployed to address and influence American legal institutions, it has since gained wider acceptance as encapsulating a more general challenge to sexist social norms regarding the treatment and role of women, not just in the US but globally. One modern day manifestation of this challenge is the 'MeToo' movement.

The third reason why critical movements are important for informal activism is that they *organise and facilitate the dissemination* of the ideas developed within the movement to nonactivists in wider society. This dissemination can take a range of forms. It might take the form of interpersonal communication, discussion, and deliberation. The British abolitionists, for instance, sought to disseminate their objections to slavery by organising lectures, meetings, and petition-signing events around the country (Appiah 2010:101-136, Anderson 2014:11-12). In large, modern societies, however, critical movements are also likely to disseminate their ideas by making use of mass media technologies. One such technology is television. Consider, for instance, the Indian soap operas *Tinka Tinka Sukh* (Happiness Lies in Small Things) and *Hum Log* (We People), both of which were designed to challenge gender inequality and promote female

empowerment and both of which led to increases in the proportions of girls attending school (Bicchieri 2017:199-202). Increasingly, critical movements are also using digital technology and the Internet to disseminate their messages. A powerful example of this is the way that the hashtag #MeToo has been used to raise awareness about the extent of the sexual harassment and abuse of women.

The relationship between critical movements as sites of innovation and critical movements as organisers and facilitators of the dissemination of ideas is reciprocal. By this, I mean that critical movements do not simply develop ideas, arguments, tactics, and slogans internally and then communicate them externally in a uni-directional fashion. Instead, the process is typically – or at least should ideally be – bi-directional, interactive, and experimental. Informal activists should develop their ideas, arguments, tactics, and slogans in a way that is responsive to how they are received by nonactivists in wider society. Jane Mansbridge captures this point well when she says that the interaction between activists and nonactivists

> combines the dynamics of a market and a conversation. In a market, entrepreneurs put forth a product, which consumers then buy or do not buy. By making this binary choice, to buy or not to buy, consumers shape what the entrepreneurs produce… Conversations, by contrast, do not depend on binary signals. An ideal conversation, like Jürgen Habermas's ideal speech situation, aims at understanding. But even a conversation has a component that works a little like a market: Each partner advances words, which the other does or does not understand, does or does not find interesting. Even a partner who does not speak can shape what the other says by nonverbally indicating understanding or confusion, interest or boredom. Nonactivists affect what activists say and think in part by being speaking partners in conversation with activists or intermediate actors and in part by responding to those offerings with understanding or confusion, interest or boredom, appropriation or rejection (Mansbridge 1999:213-4).

When this interaction goes well, informal activists develop ideas and slogans that capture the attention and win the acceptance of nonactivists, often using the mass media as the means by which to disseminate these ideas.

For an illustration, consider the word 'mansplaining'. Mansplaining is "when a man explains something to a woman that she already knows, often in a condescending tone" (Brook 2018). Inspiration for the term is typically attributed to Rebecca Solnit's 2008 essay *Men Explain Things to Me: Facts Didn't Get in Their Way*. There, Solnit recounts a story in which a man describes to her, with great authority, a book that she had actually

written and he had in fact not read. She connects this anecdote to the more general phenomenon of male overconfidence and female self-doubt (Solnit 2008). The word mansplaining began to appear on feminist blogs shortly after the publication of Solnit's essay, and use and awareness of it has increased ever since. In 2010, the *New York Times* named 'mansplainer' as one of its words of the year (Sifton and Barrett 2010). The term has been the subject of a wide range of media articles and used to describe the actions of people ranging from Mitt Romney, the republican candidate in the 2012 election for President of the USA (Cogan 2012), to the American actor Matt Damon (BBC Trending 2015). Data from the website Google Trends suggests that worldwide Google searches for the term 'mansplain' began in early 2010 and have been trending upward ever since (Google Trends 2021).

It is worth noting that the most hospitable social environment for critical movements as I have described them is one where the interests and rights which I referred to earlier – those that inform and give shape and structure to civil society – are respected and protected. If citizens do not enjoy the right to freedom of thought – if they lack access to minimally adequate stimulation and information, for instance, or if they are brainwashed and intimidated so that they do not have sufficient control over the inner workings of their minds – then they will be less likely to question prevailing social norms. If citizens do not enjoy the right to freedom of association, then, even if they have developed personal objections to the status quo, they will be less able to join with likeminded others to form critical movements. If citizens do not enjoy the right to freedom of expression, then they will be less able to communicate amongst themselves to develop their critique of the status quo, and they will be less able to disseminate this critique to wider society. Finally, if citizens do not enjoy the right to privacy, then, even if they do enjoy the right to freedom of expression, they will be less inclined to communicate their transgressive thoughts because they may fear unwelcome scrutiny and negative consequences. Hence, while civil society is not strictly necessary for critical movements – the latter have emerged and existed in societies with authoritarian and totalitarian regimes – a social environment within which civil society is acknowledged and respected facilitates and enables the formation of critical movements and means they're more likely to survive and thrive.

4.4. Persuasion as the ideal

Having highlighted the importance of critical movements, let me now turn to discuss how it is that they should go about engaging in informal activism in civil society. The first point I want to make is that persuasion is the ideal means by which critical movements should encourage nonactivists to abandon prevailing, unjust social norms and to comply with new, more just ones instead. Though persuasion is an intuitively very familiar idea, the way that I understand it is somewhat distinctive so let me unpack it here.

I define *persuasion* as a form of influence that involves inducing a person or group of people to do or believe X by: (a) bringing them to recognise and understand that there are good reasons to do or believe X, and (b) doing so in a way that does not involve subjecting them to violence (I will explain why this second condition is needed in a moment). What counts as a 'good reason' is a complex question which I cannot hope to settle in this chapter. But for present purposes I will take a 'good reason' to include any consideration that encourages the person who encounters it to endorse an act or proposition by helping that person to adhere to ideal rational standards. The plausibility of this characterisation is suggested by the fact that it well captures what many take to be the paradigmatic case of persuasion: engaging in written or verbal communication to disseminate sound or cogent arguments.[63] This activity involves advancing premises which, if believed, should lead those who encounter them to endorse a desired conclusion. Given that the arguments are sound or cogent, the premises in question are true. And because truth is one of the ideal rational standards for belief (de Sousa 1987:122, Noggle 1996:44), to induce another to endorse something by encouraging them to believe what is true is to help them to adhere to (one of the) ideal rational standards.

For an example of informal activism employing this kind of persuasion, consider the "[t]ransformative human rights deliberations" which Gerry Mackie and John LeJaune argue have led some communities in Africa to abandon FGC (2009:26). The process

[63] I don't mean to suggest that many believe persuasion to be coextensive with engaging in verbal or written communication to disseminate sound or cogent arguments; I'm sure many also take persuasion to cover less austere methods of influence. Nevertheless, deploying verbal or written communication to disseminate sound or cogent arguments does seem to me to be widely understood to be the *paradigmatic case* of persuasion. I myself take the concept of persuasion to cover more forms of influence besides this, as I will go on to explain in the text.

typically begins with an NGO convincing a small group of community members to reject FGC by educating them about their human rights, thereby creating what I would call a critical movement. The members of this newly formed critical movement then disseminate their newly acquired information to other members of their community via discussions in schools, churches, and other local institutions. Through these discussions, more and more of the community are convinced to reject FGC, until it is possible to organise a public commitment to abandon the practice. According to Mackie and LeJaune, "[i]n the first observed mass abandonments, in both Egypt and Sudan, public commitments to end FCM/C, came only *after* human rights deliberation was introduced" (Mackie and LeJaune 2009:26 italics in original).

But I also understand persuasion to include two other forms of influence that may not typically be so conceived. The first are appeals to emotion. It is tempting to think that an appeal to emotion cannot count as persuasion because an emotional response cannot count as a good reason to do or believe anything. But this is not so. Following Ronald de Sousa, I take emotions to be governed by ideal rational standards, just as beliefs are. One of the ideal rational standards for emotions is that they be caused by properties that are intelligible rationalisations of them, and which properties count as intelligible rationalisations will differ depending on the emotion in question. For an emotion to be caused by a property that could not be an intelligible rationalisation of it is for that emotion to be inappropriate and, to that extent, irrational. As de Sousa puts it:

> it is bizarre to experience intense amusement at the perfectly familiar taste of potatoes; it is morally sinister to feel euphoria at the intense suffering of an innocent person; it is altogether unintelligible – in the absence of some clever story – to be told that someone experienced excruciating remorse at the thought of having once smiled at a child (de Sousa 1987:143).

Given this, it seems to me that, so long as the emotion being elicited is appropriate to the situation, emotions can count as good reasons and emotional appeals can count as persuasion. An example of an emotional appeal that is also an instance of informal activism by persuasion on this view would be an advertising campaign that sought to influence people to cease their consumption of animal products by vividly depicting the suffering that animals often endure in the course of the making such products.

The second kind of influence which I take to be a form of persuasion, but which is not typically so understood, overlaps considerably with a form of activism which –

inspired by Clarissa Rile Hayward – we might call *epistemic disruption*. Epistemic disruption takes place when activists "withdraw cooperation from an epistemic power relationship", where the dynamics of this epistemic power relationship prior to the disruption "enable[d] motivated ignorance" (Hayward 2020:455). As an example of such an epistemic power relationship, Hayward cites the relationship between black Americans and white, liberal Americans. Typically, the dynamics of this epistemic power relationship allow white, liberal Americans to pay little to no attention to the reality of racial inequality in the US. This reality is inconsistent with the principles white, liberal Americans claim to endorse and so attending to it would be uncomfortable and would unsettle their belief in their own goodness. And so, in a less than fully conscious way, white, liberal Americans choose to direct their attention away from the reality of racial inequality and so to remain ignorant of it; this is what Hayward means by 'motivated ignorance'.

According to Hayward, when activists engage in unruly actions – such as sit-ins or interruptions of concerts – they ensure that issues such as racial injustice can no longer be ignored. White, liberal Americans must, then, attend to the fact that the principles they claim to endorse are not being realised. While Hayward claims that unruly actions which aim at epistemic disruption "are not a matter of moral suasion" (Hayward 2017:406), such actions do counts as forms of persuasion as I understand the term. This is because they aim to induce others to act – in this case to advance racial justice – by bringing them to recognise and understand that there are good reasons to act. It just so happens that, in the case of epistemic disruption, the reasons activists bring nonactivists to recognise and understand are reasons nonactivists already endorse (principles prescribing racial equality, for instance) but have chosen to ignore.[64]

[64] For what it is worth, it seems to me that Martin Luther King Jr., one of the most famous proponents and practitioners of the kind of unruly, disruptive activism Hayward defends, would agree with my characterisation of the acts he engaged in as forms of persuasion. The purpose of nonviolent direct action, King says in his *Letter From a Birmingham Jail*, is "so to dramatize the issue that it can no longer be ignored" (King 1963). This is a perfect description of what Hayward means by epistemic disruption. However, King also regarded the action he engaged in to be a form of persuasion. When speaking of nonviolent direct action in his Nobel Lecture, King said:

> We will do this peacefully, openly, cheerfully because our aim is to *persuade*. We adopt the means of nonviolence because our end is a community at peace with itself. We will try to *persuade* with our words, but if our words fail, we will try to *persuade* with our acts. We will always be willing to

It is worth saying that I do not regard every form of influence that might fall under Hayward's definition of epistemic disruption as a form of persuasion. To illustrate, consider the following. Imagine that two people are in an unequal epistemic power relationship in which the more powerful member is regularly cruel to the less powerful member. However, because the more powerful person considers themself to be a good person, they, in a less than fully conscious way, choose to ignore or downplay their cruelty. Perhaps they tell themselves that they are 'just joking around' and the less powerful member is in on and enjoys the joke. Now suppose that, as a last resort, the less powerful person slaps the more powerful person in the hope that this will jolt the more powerful person into recognising what they already know – that they are regularly cruel to the less powerful person. This slap would count as an act of epistemic disruption on Hayward's account, but, because of the violence involved, I do not regard this as an instance of persuasion.[65] The purpose of condition (b) in the definition of persuasion I offered above is to exclude this kind of case.

Why do I say that persuasion is the ideal form that informal activism in civil society should take? The reason relates back to the conception of the person that I introduced in the previous section. Recall that a concern for civil society is informed by a conception of human beings as creatures who are not merely responsive to prudential reasons but are also able to understand and be motivated by moral reasons. I described this capacity as our moral personhood. Persuasion is the form of influence that most faithfully respects our moral personhood, and, for this reason, it is the ideal means by which human beings may influence each other in civil society.

4.5. The presumption against pressure

Having introduced persuasion as the ideal form that informal activism in civil society should take, let me now contrast it with two other means by which critical movements might seek to encourage nonactivists to abandon prevailing, unjust social norms and to comply with new, more just ones instead.

> talk and seek fair compromise, but we are ready to suffer when necessary and even risk our lives to become witnesses to truth as we see it (King 1964 emphasis added).

[65] At least it is not a pure or straightforward example of persuasion. Perhaps it is a hybrid of persuasion and what I will come on to call 'price-setting'.

The first alternative means is what I will call pressure. I understand *pressure* to be a form of influence that involves inducing a person or group of people to do X by either physically forcing them to do X or by making it so that by refusing to do X they would put their fundamental interests at risk.[66] Fundamental interests are those that must be protected in order for a person to have a minimally decent life (see Miller 2007:178-185). I take them to include (but not be exhausted by) interests such as the interest in subsistence and the interest in physical security, and the interests I described as fundamental in the previous section – the interests in having and honouring conscientious moral convictions, the interest in having intimate social connections, and the interest in having control over how we present ourselves to others. Though it is worth noting that in the latter cases putting these interests at risk would involve threatening to undermine the broad ability that the interest picks out, not any particular instantiation of that ability. So, one's interest in having intimate social connections is put at risk when one is threatened with total social ostracism or solitary confinement, but not when a former friend chooses to no longer associate with one.

Prima facie, pressure should be effective as a form of informal activism by providing those who currently comply with unjust social norms with a compelling incentive to change their behaviour. Nevertheless, despite this intuitive rationale, I regard pressure as presumptively impermissible and I do so precisely because it poses a risk to others' fundamental interests. What it means for an interest to be fundamental is for it to be of such great importance that it should almost always be protected.

The one exception – the reason why pressure is *presumptively* impermissible – is self-defence, where an attacker unjustly threatens a defender's fundamental interests. For instance, consider a woman who is faced with an attacker – her brother, say – who is

[66] It seems to me that what I label pressure could just as easily be called 'coercion'. Jane Mansbridge, for instance, defines coercion in a very similar way to how I have defined pressure (Mansbridge 1996). Furthermore, Grant Lamond characterises coercion in terms of pressure when he says that "[w]hat is common to all modes of coercion, and what makes them instances of coercion, is the underlying notion of sufficient pressure being brought to bear by one person to *force* or *make* another person do as the first wills (Lamond 2000:43 italics in original). I have chosen not to speak of coercion in part because there is a disagreement in the literature about how this term should be defined and I want to avoid unnecessary terminological debates. According to the definition I subscribe to, offers (not just threats) can constitute pressure – making immediately necessary, life-saving aid conditional on a person or group of people doing as the giver of the aid wishes would be an example of a pressurising offer, for instance.

intent on killing her to restore their family's honour. It is permissible for this woman (or others acting on her behalf) to use pressure to avert the threat she faces and so to prevent her attacker from complying with the social norms underpinning the practice of 'honour killing', so long as her use of pressure adheres to the principles of permissible defensive harm discussed in the previous chapter: success, necessity, and proportionality. What the presumption against pressure rules out is the use physical force or threats to fundamental interests in scenarios that cannot be conceived of as cases of self-defence.

To illustrate, consider the recent trend of activists inflicting or threatening to inflict violence on those whom they deem to be proponents of unjust social norms. Examples include the assault of American white nationalist Richard Spencer in 2017, or the threats of violence made against feminists who resist efforts to open women-only spaces to transwomen. One way to interpret these events is as attempts to create strong incentives to deter others from promoting or complying with social norms that the activists regard as unjust by proposing to threaten the bodily integrity of those who do not do as the activists wish. In light of the presumption against pressure, I regard these forms of informal activism as impermissible.

Before I introduce the second alternative to persuasion, let me respond to two objections to what I have said so far about pressure. First, an objector might point out that, when unjust social norms become widely accepted or even tolerated, pressure deployed against marginalised groups itself becomes more likely. Hence, using pressure to deter people from promoting unjust social norms can be understood as a form of 'preventive' self- (or other-) defence (see Bray 2017).[67]

In response, while I concede that using pressure in this way will sometimes be permissible, it seems to me that the circumstances in which this will be so will be rare. To see why, consider a variation of the honour killing example introduced in the previous paragraph, in which a man was intent on killing a woman – the man's sister – so as to restore their family's honour. In that case, I suggested that it would be permissible for the

[67] Mark Bray argues that it is legitimate to use physical force to "shut down" occasions in which fascist and white supremacist ideology is espoused so as to prevent the violence that he believes eventually follows the espousal of this ideology (Bray 2017). Bray refers to this at approach as "preemptive self-defence", but it seems more appropriate to call it 'preventive self-defence'. Preemptive self-defence suggest that the threat being defended against is imminent, whereas I take preventive self-defence to also include defence against threats that are expected to occur at some point during some longer future time frame.

woman (or others acting on her behalf) to use pressure to deter the man from complying with the unjust social norm. Now imagine that this man has devised a plan kill the woman with his (and the woman's) mother, who is also intent on killing the woman (her daughter) so as to restore the family's honour. The man is known to lose his nerve in stressful, high stakes situations, so the mother's role in the plan is to watch over the ambush from a safe distance and to supply her son with necessary encouragement at the crucial moment by shouting to him that he should 'do his duty and restore his family's honour!' just before he is due to strike the fatal blow.

Now suppose that a neighbour who overheard the mother and son as they were devising their plan comes across the mother in her hiding place. The neighbour has lost her voice so cannot shout a warning to the woman who is about to be attacked, but the neighbour is able to use pressure to prevent the mother from communicating with her son. My view is that it would be permissible for the neighbour to use pressure in this way in these circumstances. This, then, is an example of a case in which it is permissible to use pressure to deter someone from espousing an unjust social norm, so as to prevent a situation in which unjust threats to people's fundamental interests are more likely.

There are, however, two features it is worth noting about this case. The first is that, even though it is the son who threatens to wield the fatal blow, the mother is also liable to defensive harm in the form of pressure. According to account I favour, one becomes liable to defensive harm by posing an unjust threat to others' fundamental interests in one of the following three ways. First, by *directly* posing the threat without further intervening agency. Second, by *jointly* posing a threat by playing an integral part in a coordinated plan to pose the threat without further intervening agency. And third, by *indirectly* posing a threat by commanding those who are subordinate to one to directly pose the threat (Haque 2017:58-66). In the example under discussion, the son is liable to defensive harm under the first criteria. However, the mother and son are both liable to defensive harm under the second criteria: they each perform an integral role in their coordinated plan to kill. In fact, the mother may also be liable to defensive harm under the third criteria. If the son is committed to obeying her as his superior, then her communication with him at the crucial moment would constitute a command.

It seems to me that many, perhaps most, people who espouse unjust social norms have not made themselves liable to defensive harm in any of these three ways. Most occasions in which unjust social norms are espoused are not themselves direct threats –

if they were, they would not need to be considered under the rubric of *preventive* self-defence – and nor are they indirect threats or constituent parts of joint threats. In the majority of cases in which people espouse unjust social norms, each one of those who does so can at most be accused of superfluously contributing to creating a situation in which unjust threats are posed by others.

To illustrate, let us assume for the sake of argument that the rule 'women-only spaces should only be accessible to people of the female sex, which does not include transwomen' is unjust. Moreover, let us assume for the sake of argument that there are a small number of people who are inspired to commit acts of violence against transwomen if they perceive very many people to endorse the rule just stated. It is extremely unlikely that any given espousal of this social norm is going to make the difference when it comes to inspiring another to commit an act of violence, but together lots of occasions in which the norm is espoused may aggregate in such a way that does inspire such violence. Hence, each espouser individually makes a superfluous contribution to this situation. I do not believe that making a superfluous contribution to an unjust threat posed by another makes one liable to defensive harm in the form of pressure. This does not mean that there is no duty not to make such contributions. Instead, what it means is that the duty not to make superfluous contributions to unjust threats posed by others is not weighty enough to justify the use of pressure to enforce that duty (see Haque 2017:70-2).

It does not follow from the fact that a person is not liable to defensive harm that defensive harm can never be used against that person. However, as I mentioned in the previous chapter, in order for the imposition of defensive harm to be permissible in this kind of case, the benefit that such an imposition yields (or the harm that it averts) must be very substantial in comparison to the harm that it causes. This brings me to the second feature of the case under discussion I would like to draw attention to: the neighbour who is in a position to engage in defensive harm has compelling evidence that the man, jointly with his mother, is about to unjustly threaten the woman's fundamental interests. The neighbour overheard the mother and son devising their plan, she has found the mother observing from safe distance, and she can see as the son lies in wait to ambush his sister. Not only does the neighbour have compelling evidence that the threat is likely to materialise, but she also has good reason to be confident that she is able to avert it. I take it that, in order for defensive harm to be permissible, the agent considering whether to engage in the harm must have sufficient evidence that an unjust threat to fundamental

interests is sufficiently likely, can be averted via defensive harm, and cannot be averted by other means (see McMahan 2016:171). I will not specify where the thresholds for 'sufficient evidence' and 'sufficiently likely' should be set, but I take it that, wherever they are set, the case currently under discussion meets these thresholds.

It seems to me that, in many, perhaps most, cases in which a person is considering whether to deploy pressure to deter the espousal of an unjust social norm so as to prevent future unjust threats from materialising, the person in question will face a great deal of uncertainty. In particular, it seems unlikely that they will have compelling evidence that: (a) the future threat that they are ultimately hoping to avert is very likely to happen; (b) presuming the future threat is likely to happen, the harm they are considering whether to impose will effectively avert the threat; and (c) the threat that they are seeking to avert cannot be averted by other means. This uncertainty is magnified even further if the person considering whether to deploy pressure also has to assess whether the harm the pressure may avert is sufficiently substantial in comparison to the harm that the pressure will involve to justify deploying pressure against non-liable targets. Given all of this uncertainty, it seems to me that the case for deploying pressure against non-liable targets so as to deter the espousal of unjust social norms will typically be very weak. It may often be the case that the evidence the person considering this course of action has is not strong enough to meet the threshold for permissibility referred to at the end of the previous paragraph.

Let me now turn to the second objection to what I have said about pressure. A critic might ask why it is presumptively impermissible for informal activists to use pressure to pursue their aims when the state and its agents use pressure all the time. Indeed, pressure is the state's *raison d'être*. The answer is that state pressure should itself be conceived of as a form of self-defence (see Renzo 2011). To see why this is so, we must return to the discussion of the previous chapter. There, we saw that, when large numbers of individuals who do not share ties of kinship or close community live in close proximity to one another in the state of nature, they tend to be drawn into mutually destructive conflict which makes life intolerably insecure. I take it that it is one of our fundamental interests to avoid this kind of conflict and insecurity.

There is, then, a sense in which individuals who live in close proximity to others in the state of nature pose a threat to each other's fundamental interests; their mere presence creates conflict and insecurity. Ordinary individuals in the state of nature

cannot end this condition of conflict and insecurity themselves because, even if all could agree on common rules of co-existence, ordinary individuals lack the power to enforce these rules and so they will quickly unravel. State pressure, however, provides those who would wish to violate common rules a compelling incentive to comply with them and those who do not wish to violate common rules the assurance that their compliance will not be exploited. In this sense, the fundamental function of state pressure is to end the condition in which individuals are posing a threat to one another's fundamental interests. In other words, state pressure is the means by which each citizen defends themselves against the threat that would have been posed by all other citizens in the absence of the state.

4.6. The permissibility of price-setting

The second alternative means of informal activism which I would like to contrast persuasion with occupies the conceptual space in between persuasion and pressure; I will call it price-setting. *Price-setting* is a form of influence that involves inducing a person or group of people to do X by proposing to attach positive consequences to them doing X or negative consequences to them not doing X. Like pressure, price-setting works by providing its targets with incentives to engage in the desired behaviour. However, because fundamental interests are not at stake, the incentives price-setting provides its targets are weaker than those provided by pressure.[68]

I do not want my use of the language of 'price-setting' to give the impression that the positive and negative consequences involved must necessarily be monetary or financial. A wide variety of situations and consequences can create the incentives that price-setting entails – from threatening to disassociate[69] from those who do not act in desired ways to offering in-kind assistance and favours to those who do – but perhaps

[68] It will be noticed that the distinction between pressure and price-setting will sometimes be blurry, in part because the answer to the question of whether a given interest is fundamental or not will be uncertain or controversial. This is to be expected. But the fact that there are difficult cases at or around the border between two proposed categories should not in itself fatally undermine the rationale for drawing a distinction between the two categories.

[69] Recall that a relatively isolated threat or decision to dissociate from someone would not put that person's fundamental interest in intimate social connection at risk, though subjecting them to total social ostracism or solitary confinement would.

the most commonly deployed are social approval and disapproval. For an illustration of how price-setting can be used as a means of informal activism in civil society, consider the recent trend of activists calling for those who endorse and propagate social norms the activists deem to be unjust to be refused a platform at university events or to lose their positions of standing and esteem in society. One way to understand these interventions is as efforts to create incentives for others to abandon the social norms the activists regard as unjust and to comply with new social norms instead.

So long as fundamental interests are not at stake, it seems to me that price-setting is a permissible means by which to pursue informal activism in civil society. Indeed, if fundamental interests were at stake, it would no longer be price-setting but would be pressure instead. One might wonder how price-setting can be permissible given that it departs from what I previously described as the ideal form that informal activism in civil society should take: persuasion. To see why price-setting is permissible, we must first recognise that, while it is respect for our ability to understand and be motivated by moral reasons that underpins civil society and that picks out persuasion as the ideal means by which we can influence each other in civil society, moral reasons are not the only kinds of reasons we can understand and be motivated by. We are also beings who can engage in instrumental reasoning and respond to prudential considerations. Indeed, we must often deploy and exploit our capacity for instrumental reasoning in order to live our lives in accordance with the conscientious moral convictions that are the end products of our moral reasoning. In part, this is because our moral convictions, goals, and life-plans are typically radically divergent, but we often depend on the cooperation of others in order to realise them. And the fact that we have radically different commitments means that we cannot assume that cooperation will be forthcoming simply because others are motivated by the same moral reasons that motivate us.

To illustrate, imagine a community that contains a small group of Christians who want to build a church but who do not have all of the skills and expertise required to do so. Suppose that there are other members of the community who do have such skills but are atheists; the religious considerations that inspire the Christians are not sufficient to motivate these other individuals to build the church. In this kind of case, the Christians might offer those who have the necessary skills some sort of in-kind assistance in the future in exchange for help now. By proposing to attach positive and negative consequences depending on whether others do or do not act in the ways that we desire,

we create incentives for cooperation in conditions in which our moral convictions are radically divergent.

It is also important to note that there is a sense in which price-setting is an inevitable – and sometimes even unintended – consequence of the exercise of the rights that establish the shape and dynamics of civil society. If people make use of their rights to freedom of thought and expression to criticise social norms that are prevalent in society, then their opposition to these norms may become widely known. This means that those who continue to endorse and comply with the social norms that are rejected by some will be aware that their behaviour is being disapproved of by other members of their society. This disapproval is also likely to manifest itself in people's associative decisions. People may make use of their right to freedom of association to associate with those who endorse and comply with the social norms they regard as just and to disassociate from those who endorse and comply with social norms that they regard as unjust.

The perception that others are disapproving of one's behaviour is something that most people find unwelcome and so seek to avoid. Hence, the fact that a given kind of behaviour is likely to generate disapproval in the minds of others can sometimes serve as an incentive for those who would have otherwise engaged in or expressed support for that behaviour to refrain from doing so. This incentive can be especially strong when the people doing the approving or disapproving occupy positions of privilege and status in society. Moreover, this holds even when disapproval is unlikely to be expressed publicly; the mere knowledge that one would be disapproved of can be enough to encourage others to amend and transform their behaviour (Brennan et al. 2012:224-5). This may be what conservatives are referring to when they complain that academic institutions, for instance, are a hostile and chilling environment for those who do not endorse or comply with 'progressive' social norms.

Nevertheless, despite the utility and inevitability of price-setting, I also want to suggest that it be regarded as a nonideal means by which to engage in informal activism in civil society. There are at least two reasons for this, one principled and one practical. The first, principled reason relates to how I characterised civil society earlier. Recall that civil society is underpinned by respect for moral personhood – the ability to understand and be motivated by moral reasons. Moreover, civil society is given its shape and structure by the fact that people have a number of interconnected fundamental interests.

In particular, we have interests in having and honouring the conscientious moral convictions that are the end products of our moral reasoning, in forming intimate associations with others who share our conscientious moral convictions, and in having control over how we present ourselves to others in part so that we are able to form and maintain intimate associations with those who share our conscientious moral convictions.

This conception of civil society does not sit easily alongside price-setting as a means by which to influence each other in civil society. To see this, consider the three paradigmatic examples of civil society institutions I mentioned earlier: religious associations, friendships, and families. Of course, these practices contain elements of price-setting: the donations that are given in church in part because of the social disapproval that not donating would generate, the favours that are done for friends in part because reciprocation is expected in the future. Nevertheless, the character of religious associations, friendships, and families would be utterly transformed if price-setting became the only, or even the dominant, means by which people influenced each other in them. If these institutions are not to be debased and lose much of their value, then at their heart there must be a conception of human beings as beings who are responding to and acting principally on the basis of moral reasons. I think this is true of civil society more generally.

The practical reason for regarding price-setting as a nonideal form of informal activism is that change secured via price-setting alone – that is, in the absence of persuasion – is inherently unstable. This is because of the phenomenon Timur Kuran calls "preference falsification". To understand this phenomenon, we must make explicit two very familiar features of human social life. The first is that a person's public preferences – the beliefs and wants they choose to express publicly – need not perfectly reflect their private preferences – that is, what they personally most prefer. The second is that human beings tend to take an interest in other people's private preferences and to respond in different ways depending on what, inferring from others' public preferences, they perceive others' private preferences to be. Responses to others' preferences can be negative – "raised eyebrows", "negative remarks", or "the denial of opportunities" – or positive – "smiles, cheers, compliments, popularity, honors, privileges, gifts, promotions, and protection" (Kuran 1995:29) – and the knowledge that different preferences are likely to elicit different patterns of responses generates incentives for individuals to

manifest whichever public preference is likely to yield the highest payoff. Preference falsification, then, is "the act of misrepresenting one's genuine wants under perceived social pressures" (Kuran 1995:3).[70]

Examples of preference falsification include the Jews living during the Spanish Inquisition who accepted baptism and lived outwardly as Christians but continued to practise Judaism in the privacy of their homes (Kuran 1995:6-7), or the gay person who seeks to conceal their sexuality by cultivating the impression that they are in romantic relationships with members of the opposite sex or perhaps even by actively discriminating against those who are openly gay (Kuran 1995:10). These examples also show how preference falsification can be self-reinforcing. The gay person who conceals their sexuality by discriminating against those who are openly gay creates incentives for other gay people to also conceal their sexuality. Similarly, during the Spanish Inquisition, many Jewish converts to Christianity participated in the persecution of practising Jews. As Kuran notes, it is possible that some of those who did so were themselves continuing to practise Judaism privately and so were persecuting Jews in order to publicly signal their own commitment to the Christian faith (Kuran 1995:62). Hence, their acts of preference falsification would have contributed to the pressure others would have felt to falsify their preferences too.

Informal activism via price-setting encourages preference falsification because, unlike persuasion, it does not seek to change others' behaviour by bringing to light moral reasons designed to target others' private preferences or personal normative beliefs. Instead, informal activism via price-setting targets public preferences. More specifically, informal activism via price-setting seeks to change others' behaviour by intervening in their choice environment so as to make it prudentially rational to publicly express whichever preference the informal activists favour. While this approach may sometimes yield widespread compliance with a social norm that a group of informal activists regard as (more) just, it may do so despite the fact that many of those who comply with the social

[70] Compare Nagel:

> To some extent it is possible to exercise collective power over people's inner lives by controlling conventions of expression, not by legal coercion but by social pressure. At its worst, this climate demands that people say what they do not believe in order to demonstrate their commitment to the right side – dishonestly being the ultimate tribute that individual pride can offer to something higher (Nagel 1998:23-4).

norm in question do not in fact endorse it. Indeed, many of those who comply with the social norm but do not endorse it may, in an effort to conceal their true preferences, actively contribute to the environment in which compliance with the social norm in question is incentivised.

This means that widespread compliance with social norms secured via price-setting alone is liable to fray and unravel. In other words, informal activism via price-setting runs the risk of creating what I have called fragile social norms. Recall (from chapter 2) that a fragile social norm is a social norm that is followed where at least one of the beliefs upon which compliance depends is false. In the case of informal activism via price-setting, the belief that is likely to be false is the belief that most other members of one's reference network have personal normative beliefs that correspond to the rule in question.

For an illustration, consider again the rise in hate crime and the open expression of white nationalist sentiment that has followed the election of Donald Trump as President of the USA 2016. It is at least plausible to interpret this series of events in the following way. Prior to 2016, informal activists had been reasonably successful in their efforts to encourage (comparatively) widespread compliance with social norms prohibiting abuse of racial minorities or proscribing overt, explicit support for white nationalist ideology. However, many of those who complied with these social norms did not in fact personally endorse them but were complying with them because of perceived social pressure to do so. That is, they were engaging in preference falsification. When Donald Trump ran for President in 2016, he aired sentiments privately harboured by many Americans and, when he won, many of those who harboured these sentiments inferred that many others harboured them too. The perceived social pressure to comply with social norms prohibiting the abuse of racial minorities and proscribing overt support for white nationalist ideology therefore lessened. And, as a result, the incidence of hate crime and of overt, explicit support for white nationalist ideology increased.

This explanation of the events that followed the election of Donald Trump is consistent with the one given in chapter 2 where I argued that the social norm proscribing abuse of racial minorities was a fragile social norm for some prior to Trump's election in 2016. The detail being added here is that preference falsification is one of the micro, individual-level mechanisms that can underpin a fragile social norm. What this suggests is that, by securing compliance with a social norm via price-setting alone,

informal activists may actually be creating fragile social norms which are vulnerable to fray in the face of events that reveal that they are sustained by widespread preference falsification. The only way to overcome this vulnerability is to attempt to change others' behaviour by targeting, not just their public preferences, but their private preferences too. That is, to engage in persuasion, not just price-setting.

4.7. Conclusion

In this chapter, I have explored how informal activism should be practised. There were two parts to my discussion. In the first, which included sections 4.1-4.2, I highlighted the limits of the state-mediation thesis – that is, the widely held assumption that ordinary citizens should discharge their duties to engage in social justice activism by encouraging the state to reform and/or enforce law and policy. The state-mediation thesis is limited because there is a protected area of social life in which the state's ability to legitimately intervene is limited. I called this area civil society and, in the second part of the chapter, which included sections 4.3-4.6, I explored how informal activism in civil society should be practised. I argued that the value that should guide the practice of informal activism in civil society is the very same value that informs civil society itself: respect for moral personhood. Having explored activism that intervenes in law and social norms, in the next chapter I will turn my attention to the third site of social justice activism I introduced in chapter 2: stereotypes.

5. Stereotypes, social environments, and social justice activism

> There is… no way out of racial injustice but through prejudice reduction.
>
> Madva 2020:228

Introduction

The previous two chapters have focused on two different sites of social justice activism. In chapter 3, I examined the ethics of efforts to remedy social injustice underpinned by law, zooming in on illegal and violent methods in particular. We have a clear idea of how injustice can be caused in this way, and, while it may be less commonly discussed, we are reasonably clear how it is that social injustice might be generated by social norms. In chapter 4, I explored how activism that seeks to change unjust social norms should be practised. Of the three social phenomena that I introduced as sites of social justice activism in chapter 2, the operation of stereotypes is the most subtle and mysterious. Recall (from chapter 2) that a stereotype is a widely distributed representation of a social group that helps those who possess it to make sense of their environment by simplifying the characteristics associated with the group being represented. As we saw in chapter 2, stereotypes can underpin social injustice, just as laws and social norms can. The goal of this chapter is to investigate the role that stereotypes can play when it comes to generating social injustice and to identify the implications of this role for the practice of social justice activism.

I have two aims in particular. The first is to highlight the importance and distinctiveness of the role that stereotypes play in generating social injustice. To this end, I will examine a recent debate between Elizabeth Anderson and Tommie Shelby. Anderson and Shelby's disagreement concerns the persistent, pervasive, multi-dimensional inequality that exists between black Americans and white Americans – which is widely regarded to be a case of serious social injustice – and in particular how to remedy this state of affairs. On the one hand, Elizabeth Anderson argues that racial

injustice in the USA[71] should be remedied via increased racial integration, and one of the means by which to achieve this goal which she advocates is state action to promote racially integrated neighbourhoods. Tommie Shelby, on the other hand, objects to Anderson's defence of state action to promote racially integrated neighbourhoods, insisting that this is not a legitimate means by which to remedy racial injustice. I will defend Anderson against Shelby's critique. More specifically, I will argue that Anderson's advocacy of racial integration becomes much more plausible once we recognise the significant and pervasive role that negative stereotypes about black Americans play in sustaining racial injustice.

My second aim is to draw out and develop two key insights generated by my analysis of the debate between Anderson and Shelby, and to use them as a platform on which to build an account of a form of activism I will call 'stereotype activism'. Stereotype activism involves attempting to remedy social injustice by changing the stereotypes that are present in a culture. The integration that Anderson recommends is a form of stereotype activism according to this definition. I draw attention to another means by which we should pursue stereotype activism: the media.

The chapter will proceed as follows. In section 5.1, I outline the disagreement between Anderson and Shelby, and defend Anderson against Shelby's critique. In section 5.2, I draw out and develop two insights from my defence of Anderson. In section 5.3, I build on these insights to develop an account of stereotype activism. Section 5.4 concludes.

5.1. Racial injustice, feasibility, and the ethics of integration

My aim in this section is to provide a very general overview of the disagreement between Anderson and Shelby regarding the ethics of racial integration, and to show, by invoking the concept of feasibility and highlighting the important role played by stereotypes, how Anderson's view can be defended against Shelby's critique.

[71] In what follows, whenever I mention 'racial injustice in the USA', 'racial injustice in the United States', or simply 'racial injustice', I will be referring specifically to black-white inequality in the United States of America.

5.1.1. Anderson and Shelby on the ethics of racial integration

In *The Imperative of Integration*, Anderson is concerned to detail, explain, and ultimately propose remedies for black-white inequality in the USA. The basic facts of the matter are relatively clear and well-understood. Black Americans are worse off than white Americans across multiple dimensions. Compared to whites, blacks have lower average life expectancy, are more likely to be in poverty, have lower incomes and less wealth, are more likely to be unemployed, do less well in education, and are vastly more likely to be imprisoned. The United States is also characterised by very high levels of black-white segregation: blacks and whites tend to live in separate neighbourhoods, go to separate schools, work in different firms, and engage in low levels of informal social interaction with each other. Anderson argues that this condition of racial segregation is one of the key factors that explains black-white inequality in the USA (Anderson 2010:23-43).

It therefore stands to reason, according to Anderson, that integration – the negation of segregation – must be part of the remedy for racial injustice. Hence, in her view, racial integration is "an imperative of justice" (Anderson 2010:112). Anderson discusses a range of integrative measures which she suggests would, under the right conditions, help to combat racial injustice. Examples of such measures include racially integrated schools, workplaces, and civic institutions such as juries. However, perhaps the most striking intervention Anderson advocates involves the state promoting racially integrated neighbourhoods by giving blacks in economically disadvantaged, segregated neighbourhoods housing vouchers to move to more advantaged, racially integrated neighbourhoods if they wish to (Anderson 2010:118-120).

Tommie Shelby offers a wide-ranging critique of Anderson's defence of state action to promote racially integrated neighbourhoods. His case against Anderson's view is founded on two observations in particular.[72] The first is that blacks who move from

[72] A third observation which Shelby considers relevant is the fact that it is by no means guaranteed that the racial integration of neighbourhoods would work out in the way Anderson suggests. As Shebly notes, "the empirical evidence is highly complex and thus difficult to assess and… empirical studies are rarely, if ever, conclusive" (Shelby 2016:72). Given this, he argues that Anderson's practical prescriptions should be more tentative and qualified. While I agree with Shelby that the prospects of success of integration are uncertain, it seems to me that some degree of uncertainty would afflict any empirically-informed practical proposal for how to remedy social justice. Hence, I will put this concern to the side. It is worth noting that Shelby barely considers the empirical question of how it is that what he considers to be an improved state of affairs

disadvantaged, predominantly black neighbourhoods to advantaged, predominantly white neighbourhoods are likely to incur various costs. These costs include not only greater exposure to hostility and racism from their new neighbours, but also reduced access to the established social networks, black institutions, and businesses that cater to black tastes and preferences.

Shelby's second observation is that, when it comes to remedying racial injustice, there is an alternative to promoting racially integrated neighbourhoods available to the state. This alternative involves removing all forms of legal exclusion and unjust economic disadvantage, and then leaving it to citizens to decide where to live. Ending unjust economic disadvantage would entail distributing socioeconomic resources equitably so as to establish fair equality of opportunity, and, in particular, taxing the affluent and using these funds to invest in and develop the places where blacks currently live. Shelby calls this alternative vision "egalitarian pluralism" (Shelby 2016:67).

Shelby's view is that, given the costs associated with Anderson's proposal, and given that there is an alternative available that does not involve these costs, black Americans are justified in resisting the state's efforts to promote racially integrated neighbourhoods and insisting that the state institute egalitarian pluralism instead. Hence, at least under current conditions, racial integration cannot be an 'imperative of justice' for blacks. Moreover, for the state to seek to encourage blacks to integrate by offering them the option of either remaining in disadvantaged, predominantly black neighbourhoods or moving to advantaged, predominantly white neighbourhoods, is to invite blacks to make a "choice under duress" (Shelby 2016:76). As Shelby puts it, "[p]olicies that seek to end unjust racial inequality by pushing, or even nudging, blacks into residential integration or that make much needed resources available only on condition that blacks are willing to integrate show a lack of respect for those they aim to assist" (Shelby 2016:75).[73]

---

is to be achieved. For an excellent discussion of both Anderson's and Shelby's failures to be sufficiently empirical, see Madva (2020).

[73] In her review of *Dark Ghettos*, Anderson interprets Shelby's critique of her own position through the lens of autonomy. More specifically, she suggests that Shelby judges integrative policies to be problematic because they interfere with ghetto residents' autonomy, an idea that Anderson dismisses as mistaken because such policies "simply expand the options for ghetto residents" (Anderson 2018:282). While Shelby does not ever appeal to autonomy himself, I think Anderson may be too quick in disposing of the thought

5.1.2. Defending Anderson against Shelby's critique

I believe that Shelby offers a serious challenge to Anderson's case for state action to promote racially integrated neighbourhoods. Nevertheless, I want to suggest that Anderson's proposal can be defended against this critique. The first step in such a defence involves recognising that, like Shelby, Anderson is in favour of a broadly egalitarian distribution of resources and opportunities. According to Anderson, social justice requires that there be an equal distribution of basic rights and liberties; that citizens be entitled to "a level of goods sufficient to enable them to participate as equals in society"; and that citizens have "fair opportunities to develop their talents to compete for positions of authority and jobs that pay more" than the minimum that is required to participate as an equal (Anderson 2012:18-9). Clearly, the USA falls far short of this ideal and a significant redistribution of resources from those people and places that are currently advantaged to those people and places that are currently disadvantaged would be required to achieve it.

So, realising the more egalitarian distribution of resources that both Anderson and Shelby favour would require a radical overhaul in the policy orientation that has held sway in the USA for at least the past few decades. It would require transforming the USA

---

that the integrative policies she favours are problematic from the perspective of autonomy. To see why this is so, note that, at least according to Joseph Raz (1986), in order to live an autonomous life, one must have, amongst other things, an adequate range of options to choose from. And having an adequate range of options to choose from requires, at a minimum, that one is not forced to make choices in which all but one non-trivial option involves sacrificing things that are necessary to and intertwined with the life one has set upon. It seems likely that at least some black residents of disadvantaged, predominantly black neighbourhoods will identify strongly with their neighbourhood and the people in it, and so will wish to remain loyal to it and them. For these people, having the choice between staying in their disadvantaged, predominantly black neighbourhood and moving out to advantaged, predominantly white neighbourhoods does not constitute an adequate range of options. Hence, while it may be true that the policies Anderson advocates do not *interfere* with ghetto residents' autonomy, they do presuppose a social environment in which blacks living in disadvantaged, segregated neighbourhoods are less able to live autonomous lives. Given that the state has a duty to provide citizens with the conditions for autonomy, policies that rely on the fact that some citizens have not been enabled to live autonomous lives can plausibly be interpreted as manifesting a certain level of *disrespect* for those citizens' autonomy. This is not to say that such policies are therefore impermissible, just that it is not obviously mistaken to insist that they are problematic from the perspective of autonomy.

from a country seemingly committed to free markets and minimal social security to a country that embraces high levels of redistribution and extensive state support for the poor and disadvantaged.[74] Moreover, the principal beneficiaries of this transformation would be black Americans, a group the members of which have been and continue to be the victims of serious social injustice. As a democracy (albeit a seriously flawed one), in order for this kind of radical change in policy orientation to happen, presumably some significant proportion of the American electorate would have to support it. It goes without saying that very many white Americans are currently opposed to such a change.

The second point to make is that, while Shelby has very little to say about what would be required to bring about a more egalitarian distribution of resources and, in particular, about how to overcome the white resistance to this kind of transformation, Anderson does provide some insights related to this issue. According to Anderson, white Americans tend to oppose policies that would help black Americans in part because white Americans tend to hold negative stereotypes about blacks. The content of these stereotypes is informed by the condition of black-white inequality that currently obtains in the USA. Hence, white Americans tend to hold stereotypes – often unconsciously – which represent black Americans as poor, lazy, incompetent, unintelligent, and prone to crime. These negative stereotypes combine with the fact of racial segregation and common cognitive biases to generate stigmatising explanations of black disadvantage, explanations rooted in failures supposedly characteristic of blacks as a group.

Anderson cites evidence which suggests that there is a link between these negative stereotypes and stigmatising explanations, on the one hand, and resistance to the kinds of policy changes that both she and Shelby favour, on the other. For instance, subscription to stigmatising explanations of black disadvantage "predict[s] opposition to government policies that are perceived to disproportionately help blacks, even when the policies are race-neutral, such as antidiscrimination laws and welfare" (Anderson 2010:61 reference removed). Similarly, "[w]hen whites' racial stereotypes are primed by pictures or stories about black criminals, or when whites are asked to evaluate black criminals, their support for highly punitive policies increases, whereas the same effect is

[74] According to James Ryerson, "Shelby is asking for no less than an overhaul of [the USA's] major political, economic and social institutions (roughly along the lines of a progressive Scandinavian country)" (Ryerson 2016).

not observed when they are exposed to stories about white criminals" (Anderson 2010:62 reference removed).

The third point to make is that Anderson proposes racial integration – not just racially integrated neighbourhoods but other forms of racial integration too – in part as a means by which to erode the negative stereotypes about blacks that whites tend to hold, and to undermine the racially stigmatising explanations of black disadvantage that these stereotypes help to generate. This proposal relies on the hypothesis that frequent, cooperative contact between groups on equal terms tends to improve intergroup relations by, amongst other things, undermining negative stereotypes and stigmatising ideas. Anderson describes this hypothesis as "one of the most widely tested claims in social psychology" and reports that "metanalyses of hundreds of studies" show strong support for it. As one piece of evidence, Anderson cites the fact that "[i]nterracial friendship appears to be a particularly potent factor that reduces prejudice", and "[t]he correlation between friendship and lower prejudice appears not to be the result of selection effects" (Anderson 2010:125 reference removed). As another, she cites the fact that the U.S. Army "is the most integrated institutions in the United States, not just by role but in the ease and comfort with which its members willingly associate across racial lines" (Anderson 2010:125).

Highlighting these three points helps us to see that the difference between Anderson's and Shelby's approaches to remedying racial injustice does not turn on the issue of whether the state should institute a more egalitarian distribution of resources; both Anderson and Shelby would agree that the state should do this. Instead, the difference between Anderson's and Shelby's approaches to remedying racial injustice relates to the question of how – by what means – the USA is to be transformed into a society that engages in significant redistribution of resources from rich to poor and, in particular, from white to black.

While Shelby largely ignores this vitally important question, Anderson, it seems to me, suggests an answer to it. Her answer begins from the observation that whites tend to hold negative stereotypes about blacks and subscribe to stigmatising explanations of black disadvantage, and that these stereotypes and stigmatising ideas are part of what explains whites' resistance to the redistributive policies that would remedy racial injustice. Given this, it seems likely that these negative stereotypes and stigmatising ideas will have to be undermined if a more egalitarian distribution of resources is to be

achieved. And the evidence suggests that racial integration is a highly effective means by which to undermine negative stereotypes and stigmatising ideas. Hence, we should view racial integration – including the racial integration of neighbourhoods – in part as the means by which the egalitarian distribution of resources is achieved; the former is a precondition for the latter.

5.1.3. Feasibility, soft constraints, and dynamic duties

One way to better understand my interpretation of Anderson's view is to appeal to the concept of feasibility and the various ideas associated with it. Appealing to these ideas also furnishes us with a powerful vocabulary with which to articulate the dispute between Anderson and Shelby, and to underline the plausibility of Anderson's view.

Following Pablo Gilabert, I take *feasibility* to relate to an agent's power or ability to bring about a given outcome in a given set of circumstances. Feasibility claims can be binary or scalar. The former are used to rule out certain outcomes and typically refer to considerations such as laws of logic and nature to so do. So, a group of labour activists striking to secure a salary rise "cannot succeed if they aim at securing for all workers in a nation at time *t1* a salary higher than the average salary for workers in the nation at *t1*" (Gilabert 2017:99 italics in original). Gilabert describes the considerations that are invoked to ground binary feasibility claims as imposing "hard constraints". Scalar feasibility claims, on the other hand, are used to determine the probability of success of the pursuit of a given outcome and these kinds of feasibility claims invoke economic, political, and cultural considerations to make this assessment. For instance, "[l]abor activists are more likely to unionize workers in a country where there is a strong solidaristic culture than in a country where competitive individualism is rampant even if they are not, strictly speaking, incapable of reaching high levels of unionization in either context" (Gilabert 2017:99). Because they are not inviolable, Gilabert refers to the considerations that are invoked to ground scalar feasibility claims as imposing "soft constraints".

One implication of an outcome being ruled out by hard constraints and so infeasible in a binary sense is that any putative duties to bring about this outcome are rendered null and void. This follows from the well-known and widely accepted 'ought implies can' principle. The nullification of duties is not a necessary implication of an outcome being subject to soft constraints and so less feasible in a scalar sense, however.

Instead, Gilabert argues that when agents face soft constraints, they may have what he calls "dynamic duties". According to Gilabert,

> [u]nlike normal duties, dynamic duties are not focused on achieving certain desirable outcomes within current circumstances. Their point is to change those circumstances so that certain desirable outcomes become achievable (or more achievable). Thus, dynamic duties direct a change, often an expansion, of an agent's power to bring about certain outcomes (Gibabert 2017:119).

With these conceptual tools in hand, we can say that Anderson regards the fact that whites tend to hold negative stereotypes about blacks and tend to subscribe to stigmatising explanations of black disadvantage as a soft constraint which makes the more egalitarian distribution of resources that both she and Shelby favour much less feasible (in a scalar sense) than it might otherwise be. Racial integration is necessary in part because it helps to undermine the stereotypes and stigmatising ideas that impede the realisation of racial justice. Hence, integration – including the promotion of racially integrated neighbourhoods – should be seen as a dynamic duty: a duty that ultimately aims at making it so that some desirable state of affairs is more likely to be achieved in the future. The desirable state of affairs in this case is the more equal distribution of resources and opportunities across society.[75]

What this suggests is that, while Shelby is correct to point out that egalitarian pluralism is in some sense available as a remedy to racial injustice, this largely misses Anderson's point. Shelby may be perfectly right that a more egalitarian distribution of resources is available insofar as it is not ruled out by hard constraints or infeasible in the binary sense. But Anderson's point is that pursuit of an egalitarian distribution of resources is seriously impeded – that is, is made much less feasible in the scalar sense – by the soft constraints that are negative stereotypes and stigmatising ideas, and that these soft constraints can be overcome via racial integration. Effectively challenging

[75] Ben Laurence invokes the concept of feasibility to interpret Anderson in a somewhat similar way, but for slightly different purposes (Laurence 2020:368-9). Shelby seems to suggest that, because the negative stereotypes and stigmatising ideas whites tend to hold about blacks are unjust, to invoke the concept of feasibility in this context "entail[s] capitulating to injustice" (Shelby 2016:74). However, it is not clear to me why we should characterise an attempt to devise an empirically informed strategy for remedying social injustice in this way. Quite the opposite, in fact.

Anderson's point would involve presenting an alternative account of the means by which a more egalitarian distribution of resources is to be secured, and arguing that this alternative means is in fact more feasible (in a scalar sense) than Anderson's proposal. Shelby does not offer such an account.

Shelby might say that the costs of moving from disadvantaged, predominantly black neighbourhoods to advantaged, predominantly white neighbourhoods are often going to be too high for blacks to have a duty to bear, so integration cannot be an 'imperative of justice' for blacks. And I think he would be right about that. But it doesn't follow that *others* – i.e., affluent white citizens and the state and its officials – lack a duty to engage in and promote racial integration, including the racial integration of neighbourhoods. Moreover, from where we are, all options involve disadvantaged black Americans bearing significant burdens: if they stay in their neighbourhoods, they must endure the unequal distribution of resources and opportunities that that entails; but if they move to more advantaged, predominantly white neighbourhoods, they must endure the alienation and hostility that may be associated with that.

In fact, it seems that, even in the unlikely event that egalitarian pluralism was to be instituted now, before any significant increase in racial integration was to take place, this option would also involve blacks bearing costs. This is because the radical development of currently disadvantaged, predominantly black neighbourhoods that Shelby advocates is likely to require deep interventions into these neighbourhoods and considerable and sustained involvement from outside actors. For blacks living in these neighbourhoods, the alienation caused by these deep interventions and the hostility that may result from increased interactions with outside actors may not be all that dissimilar from what they would experience if they chose to move to more advantaged, predominantly white neighbourhoods (see Poama 2020:134-5).

5.2. Stereotypes, social environments, and change

In the previous section, I defended Anderson's claim that integration must be part of the remedy for racial injustice, by emphasising the role that certain stereotypes play in making it so that remedying racial injustice is less feasible than it might otherwise be and highlighting the role the integration can play in undermining these stereotypes. My aim in this section is to draw out and develop two key insights which I think we should take

from the previous section. In the following section, I use these insights as a platform on which to build an account of a form of activism I call stereotype activism.

5.2.1. Stereotypes are embedded in social environments and affect social behaviour

There are two closely related lessons that I think we should take from the discussion of the previous section. I will discuss the first lesson in this subsection and the second in subsection 5.2.2. The first lesson follows from the fact that the condition of black-white inequality in the USA provides the content of the stereotypes about blacks that whites tend to hold and that holding these stereotypes is associated with lower levels of support for policies that would remedy racial injustice. What this suggests is that *stereotypes are deeply embedded in the social environment and can generate patterns of behaviour which underpin social injustice*.

This insight is also reflected in recent research by B. Keith Payne, Heidi A. Valetich, and Kristjen B. Lundberg into the phenomenon commonly known as 'implicit bias'. Recall (from chapter 2) that implicit bias refers to the extent to which an individual has imbibed a given stereotype, regardless of whether they identify with or endorse that stereotype, as measured by tools such as the "Implicit Association Test" or IAT (see Greenwald et al. 1998).

The aim of Payne and colleagues' research is to resolve three puzzles they identify in the literature on implicit bias. The first is that, while average, group-level performance on implicit measures such as the IAT is robust and stable, individuals' performance on these measures fluctuates considerably over time. Indeed, "the temporal stability of these biases is so low that the same person tested 1 month apart is unlikely to show similar levels of bias" (Payne et al. 2017:233). The second – which follows from the first – relates to the fact that young people perform similarly to adults on implicit measures and that this fact is often taken to show that implicit biases are learnt early and fairly permanent. "Yet, if one's biases are not stable across a month, how can they be stable across a lifetime?" (Payne et al. 2017:234). The third puzzle is that, while an individual's performance on an implicit measure predicts behaviour only weakly, performance on implicit measures is strongly correlated with discriminatory behaviour or disparate outcomes when it is aggregated to the group level. As Payne et al. report,

> in one study, countries with stronger average associations between males and science had greater gender-based achievement gaps in eighth-grade science and math scores… In another study, metropolitan regions in the United States with higher levels of implicit race bias were found to have greater racial disparities in shootings of citizens by police… In France, city-level implicit bias towards Arab (vs. French) individuals was associated with lower participation rates in city marches aimed at national unity in the wake of an Islamic terrorist attack (Payne et al. 2017:234 references removed).

Payne et al. attempt to resolve these puzzles by appealing to two key ideas. The first is "concept accessibility", which refers to "the likelihood that a thought, evaluation, stereotype, trait, or other piece of information will be retrieved for use" (Payne et al. 2017:235). A concept is more accessible when it is used often or has been used recently, or when a related concept is being used. And concepts come to be related by being regularly encountered together. "For example, the more frequently and recently one has encountered stereotypic portrayals of Black men as criminals, or women as a fragile… the more easily accessible these stereotypic attributes will be when one thinks about the social category" (Payne et al. 2017:235). It is widely accepted, Payne et al. suggest, that implicit bias is linked to concept accessibility.

The second key idea is "the bias of crowds". The central observation that underpins this idea is that, while individual people in a given society will inevitably have vastly different thoughts and encounters, they will also share knowledge about the stereotypes and prejudices of their culture. Payne et al. argue that implicit bias should not be principally understood as a feature of individuals' minds, but instead in terms of these widely and consistently accessible stereotypes. As they put it: "it is more accurate to consider implicit bias as a social phenomenon that passes through the minds of individuals but exists in the situations they inhabit" (Payne et al. 2017:236).

Together, concept accessibility and the bias of crowds help to resolve the puzzles Payne et al. identify. The combination of robust average, group-level bias and low temporal stability of bias in individuals is explained as follows. People's divergent and constantly changing thoughts and encounters will mean that concept accessibility will vary considerably across populations and for any given individual across time. But those stereotypes that are stable and widely shared within the culture will be consistently accessible for all members of society. When concept accessibility is aggregated to the group level, variation in accessibility will be randomly distributed across the individuals

and so averaged away, leaving shared stereotypes to emerge as the central group-level tendency. Average implicit bias scores reflect the consistent accessibility of stereotypes at the group level.

The combination of young people showing the same level of bias as adults and low levels of temporal stability is explained by the fact that "[w]hen children and adults are sampled from the same times and places, they reflect the same stereotypes and biases that inhabit that context… they are cued by the same environments and therefore have the same concepts accessible, on average, at any given moment" (Payne et al. 2017:238).

Finally, the fact that individual performance on implicit measures is only weakly predictive of individual behaviour but aggregated performance is strongly predictive of group-level discrimination reflects the fact that implicit bias is principally a feature of social environments. The stereotypes that are widely shared in different cultures will vary. Moreover, societies will be differently organised and arranged such that different stereotypes will be frequently accessible. For example, in a society characterised by the kind of inequality Anderson describes – where black citizens are more likely to be poor, criminalised, have lower educational prospects, and be employed in lower status jobs, for example – negative stereotypes about black people will, on average, be more frequently encountered than in a more equal society, and so will be more accessible too.

5.2.2. Eroding stereotypes will often require changing the social environment

Let me now turn to the second lesson to take from the discussion of section 5.1. This lesson follows from the thought that increased interaction on terms of equality between black and white Americans is a means by which to thwart the negative stereotypes about blacks that whites tend to hold. What this suggests is that, *in order to erode a stereotype or at least undermine its effect on behaviour, it will often be necessary to change the social environment within which the stereotype is embedded*. As with the previous lesson, we can draw on research on implicit bias to develop this insight.

Consider the following study into the effects of two naturally occurring social environments on the implicit biases of those who inhabit those environments. Nilanjan Dasgupta and Shaki Asgari investigated whether attending a coeducational college or a women's college would have an impact on the extent to which women harbour stereotypes that represent women as occupying supportive as opposed to leadership roles. Coeducational colleges and women's colleges differ insofar as the latter contain a

higher proportion of women who occupy counter-stereotypical leadership positions (e.g., as faculty) than the former.

Dasgupta and Asgari tracked a group of women who had chosen to attend a coeducational college and a group of women who had chosen to attend a women's college. They found that, while the women who attended both types of college manifested implicit bias in favour of stereotypical representations of women (i.e., as having supportive as opposed to leadership qualities) during their first semester at college, those who attended the women's college did not express such implicit biases a year later. By contrast, implicit bias in favour of stereotypical representations of women actually increased amongst those women who attended the coeducational college.

Moreover, Dasgupta and Asgari found that these changes in implicit bias were mediated by the frequency with which students at each college were exposed to female faculty (i.e., in leadership positions). That is, the more students at either college encountered female faculty in the classroom, the less they manifested implicit bias in favour of stereotypical representations of women as occupying supportive as opposed to leadership roles (Dasgupta and Asgar 2004). It seems, then, that one way to reduce the prevalence of stereotypes in a population is to change the social environment that population inhabits so that the representation that constitutes the stereotype in question is less prominently manifested in that social environment.

It is crucial to note, however, that there are more ways to change a social environment besides changing the people in it or the positions those people occupy. Another way to change a social environment is to change the public representations – e.g., the symbols, texts, and utterances – that characterise that environment, and there is considerable evidence that such changes can also have an impact on the stereotypes held by inhabitants of a given environment. For instance, in another study that was part of the same research series as the study just described, Dasgupta and Asgari found that exposing women to pictures and descriptions of famous women in high-profile leadership positions (e.g., US Supreme Court Justice Ruth Bader Ginsburg) moderates the expression of implicit bias in favour of stereotypical representations of women (Dasgupta and Asgari 2004).

Similarly, in research conducted with Anthony G. Greenwald, Dasgupta has found that exposing people to pictures and descriptions of admired African Americans (e.g., civil rights leader Martin Luther King Jr.) and disliked white Americans (e.g., serial killer

Charles Manson) has the effect of decreasing the manifestation of implicit bias in favour of negative stereotypes about African Americans (Dasgupta and Greenwald 2001). And the effects of public representations on the expression of implicit bias are not limited to pictures or descriptions of famous exemplars. Studies have found that exposing people to unfamiliar pictures or movie clips that depict African Americans in a positive context (e.g., at a family barbeque or in a church) lead to decreases in the expression of implicit bias in favour of negative stereotypes of African Americans (Wittenbrink et al. 2001).

It is not just performance on implicit measures that public representations affect; they have an impact on other behaviour too. Consider the results of a laboratory experiment known as the 'First Person Shooter Task' (FPST). This experiment involves participants playing a videogame designed to simulate encounters with potentially hostile targets. Participants are asked to make split-second decisions about whether or not to shoot targets, based on whether the target is armed ('shoot') or not ('don't shoot'). The targets in the game are either young black men or young white men, and they can either be carrying a gun or some other object (such as camera or a wallet). Studies consistently show that, despite the fact race is irrelevant to the task, participants in this experiment are biased towards shooting black as opposed to white targets (Correll et al. 2007).

Joshua Correll and his colleagues conducted a version of the experiment in which participants are asked to read an article about a violent crime before performing the task. They found that those who read an article in which the suspects are black are biased towards shooting black targets compared to white targets, even though both black and white targets in the videogame are equally likely to be armed. Conversely, participants who read the article in which the suspects of the violent crime are white show no evidence of bias (Correll et al. 2007).[76]

In another iteration of the experiment, there is a preliminary round where participants either play a version of the videogame in which black targets are more likely to be armed than white targets or vice versa, before then participating in a second round where they play a version of the videogame in which both black and white targets are equally likely to be armed. Participants who are exposed to the preliminary version of the

[76] It is worth noting that bias is eliminated in this latter condition because participants' propensity to shoot white targets increases, not because their propensity to shoot black targets decreases.

videogame in which black targets are more likely to be armed, show a greater bias for shooting black targets in the second round of the experiment, compared to those who play the version of the videogame in which white targets are more likely to be armed. Participants who are exposed to the preliminary version of the videogame in which black targets are more likely to be armed are also more likely to exhibit implicit bias in favour of the stereotypical association between blacks and danger.

The authors of these studies argue that reading a newspaper article about a violent crime committed by a black person, or being exposed to a videogame in which blacks are more likely to be armed, activated and reinforced the stereotypical association between blacks and danger in the minds of participants, thereby increasing the likelihood that they will shoot black as opposed to white targets in split-second decision-making scenarios (Correll et al. 2007).

5.3. Stereotype activism, the media, and manipulation

So far in this chapter, I have defended Anderson's advocacy of state action to promote racially integrated neighbourhoods from Shelby's critique, and drawn out and developed two key insights that my defence of Anderson generated. The first is that stereotypes are deeply embedded in social environments and can generate patterns of behaviour which can underpin social injustice. The second, related insight is that changing stereotypes will often involve changing the social environments within which they are embedded, including changing the public representations which characterise those environments. Together, these insights suggest that it will sometimes be necessary to remedy social injustice by changing the stereotypes that a given population harbours, and that one way to do this is to change the social environment that the population inhabits. Let us call any attempt to remedy social injustice by changing the stereotypes that are present in a culture *stereotype activism*.

The programme of racial integration that Anderson advocates is a form of stereotype activism, according to this definition. Recall that Anderson proposes racial integration in part as a means by which to erode the negative stereotypes about blacks which impede the realisation of social justice. Anderson's defence of racial integration also highlights the complex relation stereotype activism sometimes stands in with respect to other forms of activism. Take what I have called formal activism – that is, attempts to remedy social injustice via changes to law and policy. Many of the

interventions Anderson proposes involve using law and policy to change the social environment citizens inhabit so as to transform the stereotypes those citizens harbour. This is true of her account of state action to promote racially integrated neighbourhoods, for instance. However, in addition to being promoted by law and policy, racially integrated neighbourhoods are also a means by which to secure further changes to law and policy – in particular, policies designed to bring about a more egalitarian distribution of resources and opportunities. What this shows is that different forms of activism can be combined to create a medium- to long-term strategy whereby a given intervention is intended to trigger a sequence of interventions that will hopefully result in the ultimately desired state of affairs.

Now, just because law and policy can and should be used as a means by which to pursue stereotype activism, this does not mean that law and policy should be the only means by which to pursue stereotype activism. There are at least two reasons for this. The first is that the very stereotypes which make stereotype activism necessary may also make the implementation of the kind of laws and policies which would constitute stereotype activism unlikely. For instance, it is possible that the negative stereotypes about blacks and stigmatising ideas about black disadvantage which impose soft constraints on the ambitious goal of radically transforming American society may also impose soft constraints on the less ambitious goal of creating racially integrated neighbourhoods.

Second, when it comes to stereotype activism, there are likely to be countervailing forces at work for law and policy to contend with. Sticking with the case of racial injustice in the USA, consider the fact that the American news media seems to overrepresent African Americans in its depictions of those who are poor or perpetrators of crime (The Opportunity Agenda 2011, Ghandnoosh 2014). For example, one study which examined stories on poverty in American newsmagazines over a five-year period found that African Americans were depicted as poor people in these stories at over twice their true proportion (Gilens 1996). Another study analysed the content of news programmes aired in Los Angeles and Orange Counties and found that blacks were much more likely to be portrayed as perpetrators of crime than to be arrested according to official crime reports for the region (Dixon and Linz 2000). Being exposed to these kinds of public representations is likely to activate and reinforce the negative stereotypes white Americans tend to hold about black people.

One might reply that it is possible to significantly reduce or perhaps even eliminate the distribution of injustice generating negative stereotypes via law and policy. We might seek to regulate how public representations about different groups in society are propagated by the media, for example. The problem with this, however, is that it seems very unlikely that any such regulatory regime which was robust enough to be effective would also be consistent with the right to freedom of expression, which was discussed and defended in the previous chapter. This seems especially obvious once we recognise the distinction between the news and the non-news media. The *news media* arguably has a responsibility to accurately reflect the world as it is, and it is not implausible to think that the state may play some kind of role in ensuring that this responsibility is discharged, though what precise role it plays will of course be a highly context-specific matter. There is, however, no expectation that the *non-news media* should accurately reflect the world as it is. Indeed, such an expectation would be inconsistent with the distinctive role that the non-news media plays in human social life: to entertain and inspire, to spark our imaginations and make us wonder about both the possible and the seemingly impossible. It seems, then, that law and policy will be limited in the extent to which they are able to legitimately regulate the propagation of public representations.

Given these limitations, I want to suggest that stereotype activism should be pursued not just via law and policy but by other means too. For an example of how stereotype activism might be pursued via means other than law and policy, recall the previous section's discussion of the impact of women's colleges on the stereotypes harboured by those who attend these colleges. What this suggests is that educational institutions could serve as an important site of stereotype activism, by providing a setting within which unjustly stereotyped groups are represented in counter-stereotypical ways. Religious associations are another potential vehicle for the pursuit of stereotype activism. Religious institutions often reflect common stereotypes – about men and leadership, for instance – and they play a central, formative role in the lives of many people. Given this, it is not implausible to think that changing how different groups are represented within religious associations might lead to changes in the stereotypes harboured by the members of these associations.

The particular means of pursuing stereotype activism which I want to focus on in the remainder of this chapter, however, is the media. The reason for this focus should be

obvious enough. Stereotype activism involves attempting to remedy social injustice by changing the widely distributed representations of social groups – that is, the stereotypes – present in a given population, and the mass media is one of the principal vehicles through which representations become widely distributed. The suggestions I want to make utilise the distinction just introduced between the news and non-news media. Regarding the former, my first proposal is that producers of news media products should monitor how different social groups are depicted in their coverage of different social issues and ensure that these portrayals accurately reflect the reality of the situation that obtains in the relevant location (see Ghandnoosh 2014:36-7). So, instead of overrepresenting African Americans in its depictions of the poor or those who have perpetrated crimes, the news media should represent all social groups in line with how they are affected by these issues in reality.

However, given that the social reality in a given time and place will often bear the scars of a history characterised by significant injustice, it is not enough to simply reflect social reality. This points towards a second proposal. In addition to accurately reflecting social reality, producers of news media products should situate that reality within its proper context whenever possible (see Ghandnoosh 2014:37). So, instead of simply noting that blacks tend to do less well in education than whites, news stories and reports might also mention that the schools that white children attend tend to be better funded than the schools that black children attend.

The third proposal that I want to make relates not to the news media, but to the non-news media. More specifically, I want to suggest that an additional form of stereotype activism would be for producers of non-news media products such as films, novels, comic books, television shows, and videogames to produce such products so that they contain counter-stereotypical representations of unjustly stereotyped groups. The rationale for this proposal is based in the research cited earlier, which suggests that encountering such counter-stereotypical portrayals can moderate implicit bias. This proposal may trigger the worry that media products which are designed to contain counter-stereotypical portrayals of unjustly stereotyped groups would be vapid or mawkish, but this need not be the case.

For an illustration, take the stereotype of African American men as poor, violent, aggressive, misogynistic criminals, and consider the Oscar-winning film *Moonlight* in this context. *Moonlight* charts three stages – youth, adolescence, and early adulthood – in the

life of a black, gay, male character named Chiron. The film has an all-black cast and depicts many of the themes and issues that constitute the stereotypical portrayal of African American men in popular culture – poverty, violence, aggression, criminality, and disfunction. But it presents these themes alongside others – awkwardness, tenderness, sensitivity, longing, care, and hope. In so doing, *Moonlight* offers a beautiful and complicated – and ultimately more realistic and human – portrayal of African American men. One way to interpret the making of a film like *Moonlight* is as a form of stereotype activism – as a contribution to the effort to challenge and undermine negative stereotypes about African American men that are often promoted and recycled in the media.

5.3.1. Addressing the manipulation worry

Before concluding, I must address an objection to stereotype activism that may be forthcoming. The concern is that, by attempting to change people's behaviour via exposure to counter-stereotypical representations of unjustly stereotyped groups, stereotype activism constitutes a form of manipulation and is objectionable for this reason. Let us call this *the manipulation worry*. In order to adequately respond to this objection, we must first understand the basis for the charge that stereotype activism is manipulative.

One reason why stereotype activism might be considered manipulative is that it attempts to influence others in a way that does not involve deploying sound or cogent arguments. Instead, stereotype activism seeks to make it so that negative stereotypes of unjustly stereotyped groups are less widespread within society and so less accessible to any given member of society at any given time. Put simply, stereotype activism works by adjusting *salience*. It seeks to induce people to change their behaviour, not by offering people true premises designed to lead them to believe that they should change their behaviour, but by making it so that the stereotype that is generating the behaviour in question occupies less of people's (conscious or unconscious) attention. It might be thought that, pressure[77] aside, any attempt to influence another via means that do not

[77] As I mentioned in the previous chapter, what I am calling 'pressure' could quite easily be labelled 'coercion'. I chose not to use the latter term so as to avoid unnecessary terminological disputes and I continue to use the former term for consistency.

involve deploying sound or cogent arguments is *ipso facto* manipulative, perhaps because such attempts to influence fail to engage with their targets' capacities for reason.

The problem with this view, however, is that it appears to generate seriously counterintuitive judgements. Consider the following hypothetical case provided by Anne Barnhill:

> *Embezzlement*: Janice has embezzled money from the company she works for. Janice's father, Mike, finds out. Over the course of a weekend together, Mike repeatedly says things like, 'I didn't raise you to be a thief' and 'You should return the money.' This makes Janice feel very guilty and [as a result] she returns the money (Barnhill 2014:55-6 italics in original).

According to the view that any form of non-pressurising influence that does not involve deploying sound or cogent arguments is *ipso facto* manipulative, Mike has manipulated Janice. But this seems to me (and Barnhill) to be the wrong result. Mike has influenced Janice by appealing to her emotions, but he has not manipulated her. What this shows is that there are forms of non-pressurising influence that do not deploy sound or cogent arguments and yet are not manipulative.

What's more, in the previous chapter I argued that, while many may take the deployment of sound or cogent arguments to the paradigm case of persuasion, we should understand this term to encompass any nonviolent means of influence that involves bringing others to recognise and understand that there are good reasons for an act or belief. Moreover, I suggested that a 'good reason' is any consideration that encourages the person who encounters it to endorse an act or proposition, by helping that person to adhere to ideal rational standards. What this means is that (non-pressurising) attempts to influence need not deploy sound or cogent arguments in order to engage with a target's capacity for reason.

It seems, then, that manipulation should not be identified with non-pressurising influence that fails to deploy sound or cogent arguments. Yet we need an account of manipulation if we are to evaluate the claim that stereotype activism is manipulative. Luckily, I think that the brief discussion of appropriate emotional appeals in the previous chapter points in the direction of such an account. In chapter 4, I suggested that emotions are governed by various ideal rational standards. But it is not just emotions that must adhere to certain ideal rational standards; so too must any mental state, including beliefs and desires. According to the account I favour, manipulation just is influence that

attempts to induce a target to fall short of the ideal rational standards for mental states such as beliefs, desires, and emotions, where these standards apply not just to the mental states induced but also to the process by which they are acquired (Noggle 1996, 2018:167-8; Barnhill 2014; Coons and Weber 2014:11-12; Hanna 2015:630-1).[78]

One advantage of this definition is that it is able to distinguish between manipulative appeals to emotion on the one hand, and appeals to emotion that are not manipulative, such as the case of *Embezzlement* just discussed, on the other. For an example of a manipulative appeal to emotion consider the following. One of the ideal rational standards for emotions is that the beliefs they are based on should be true (Noggle 1996:46). But sometimes people seek to induce emotion even though the belief that would make the emotion appropriate is false. This is one of the tactics Iago from Shakespeare's play *Othello* uses. As Robert Noggle notes, "[w]hen Iago incites jealously in Othello, he gets Othello to have an emotion that is inappropriate, since the belief that would make it appropriate, namely that Desdemona is unfaithful, is false" (Noggle 1996:46).

How should we interpret the manipulation worry in light of this definition of manipulation? I think there are at least two possibilities. According to the first interpretation, stereotype activism is manipulative because it involves an attempt to induce people to fall short of the ideal rational standard that should govern salience. This interpretation of the manipulation worry does not seem very plausible. The reason is that the ideal rational standard that should govern salience is "for things to be salient in proportion to their actual importance" (Noggle 2018:166). Far from attempting to induce people to stray from this standard, stereotype activism actually seeks to move people from a situation in which they are falling short of the standard to a situation in which they are closer to it.

[78] This account manipulation raises the question: from whose perspective are we to judge whether a mental state falls short of ideal rational standards? As should be clear from the discussion to follow, I adopt an objective view according to which manipulation is influence that induces its targets to fall short of what are *in fact* the ideal rational standards for mental states. Robert Noggle defends an alternative view according to which manipulation is influence that induces its targets to fall short of what *the influencer takes to be* the ideal rational standards for mental states (Noggle 1996:47-8). Jason Hanna offers a persuasive critique of the influencer-relative view and defence of the objective view (Hanna 2015:632-6).

In societies in which negative stereotypes of unjustly stereotyped groups are generating unjust patterns of behaviour, people may fall short of the standard for salience in one or both of the following two ways. First, the offending stereotypes may have so captured the attention of the population that many actually overestimate the degree to which the stereotypes are manifested in reality. This seems to have happened with respect to white Americans' impressions of the proportions of black Americans who are in poverty, for instance (Gilens 1996). When people fail to attend to the *empirical importance* of a stereotype, we might say that they fall short of the standard for *empirical salience*.

Second, even if the population correctly judges the extent to which a given stereotype is manifested in reality – that is, even if a population adheres to the rational standard for empirical salience – if that stereotype is generating social injustice then the stereotype ought to be changed. Insofar as an unjust but 'accurate' stereotype is actively impeding the remedying of social injustice, then to attend to that stereotype is to give it outsized *normative importance* and hence to fall short of the rational standard for *normative salience*. To expose people to counter-stereotypical representations of unjustly stereotyped groups is to induce them to shift their attention to a representation of a social group that they should, as a matter of morality, be taking more notice of. Hence, I submit that, when stereotype activism is called for, it actually helps people to conform to the ideal rational standard for normative salience.

According to a second interpretation of the manipulation worry, stereotype activism is manipulative not because the mental state it seeks to induce itself falls short of ideal rational standards, but because this mental state has been acquired via nonideal means. In particular, it seems that cultural activism seeks to induce mental states via means that are non-transparent – that is, via means about which the targets of the influence are unaware.

This interpretation of the manipulation worry assumes that it is an ideal rational standard for the acquiring of mental states that the process by which mental states are acquired be transparent to those who are acquiring the mental states. But it is worth distinguishing between two senses in which a means of influence might be non-transparent. According to Jason Hanna, "[w]e can distinguish between (a) a person's lack of awareness *of* some particular thing that is in fact influencing him, and (b) a person's lack of awareness *that* some particular thing is influencing him" (Hanna 2015:642 italics

in original). An example of a form of influence characterised by (a) is subliminal messaging, since subliminal messaging works (supposedly – let us assume that it works) by influencing its targets in a such a way that they are unaware *of* any intervention, let alone any potentially influential intervention. When it comes to influence characterised by (b), however, while targets are aware *of* an intervention, they are unaware *that* this intervention is influencing them. Hanna suggest that "[t]he lack of awareness described by (b)… is characteristic of much ordinary human interaction." He continues:

> consumers are often unaware of the ways in which they are influenced by the appearance of a company's logo, but it seems implausible to conclude that a company behaves objectionably unless it tells consumers its reasons for adopting a particular design. Likewise, John F. Kennedy surely was not morally required to tell viewers of the 1960 presidential debate that they might be influenced by his famous blue suit, even if he wore the suit to impress voters. If we wish to condemn subliminal messaging but not these other forms of influence, we should probably conclude that the lack of awareness described by (a) is more troubling than the lack of awareness described by (b) (Hanna 2015:642).

The basic insight underlying Hanna's remarks seems to me to be correct and important. We should reject the view that, in order to avoid the manipulation worry, any attempt to influence must not only ensure that its targets are aware of an intervention but must also ensure that they are aware that a given intervention is influencing them. Such a view would suggest that whole swathes of human behaviour – sitting up straight, looking someone in the eye, remembering someone's name – must, when engaged in, be explicitly flagged as means of influence or else be denounced as manipulative. This is a highly counterintuitive implication. Instead, so long as the targets of a given form of influence are aware of the interventions that are influencing them, and the influence seeks to induce a mental state that meets its ideal rational standards, then it seems to me that the form of influence is not manipulative and so is not vulnerable to the manipulation worry. The reason why such a form of influence would not be vulnerable to the manipulation worry is that it would not be an instance of manipulation.

Ideally, all forms of stereotype activism would be designed so that they are not manipulative, but there is nothing in the nature of stereotype activism that guarantees that this will be so. In particular, though I argued that stereotype activism is by definition designed to ensure that its targets conform to the ideal rational standards for salience,

one might imagine a form of stereotype activism – subliminal messaging for social justice, say – that was deployed in such a way that its targets were not merely unaware *that* a given intervention was influencing them but were also unaware *of* any intervention in the first place. While I believe stereotype activism that took this form would be manipulative and would be *pro tanto* wrong for this reason, I see no reason to think that it must be deemed categorically impermissible. One might imagine an – admittedly unrealistic – scenario in which subliminal messaging for social justice was simply the only way to remedy some severe and intractable instance of social injustice. I will not pursue this line of thought any further here, but it is worth noting that pressure is – like manipulation – also *pro tanto* wrong but many agree that it too can be a permissible means by which to pursue social justice in the right circumstances.

5.4. Conclusion

Of the three social phenomena that I introduced as sites of social justice activism in chapter 2, the operation of stereotypes is the most subtle and mysterious. My aim in this chapter has been to investigate the role that stereotypes play when it comes to generating social injustice and to identify the implications of this role for the practice of social justice activism. I started by examining a recent debate between Elizabeth Anderson and Tommie Shelby about how to remedy racial injustice in the USA. I defended Anderson's position in this debate, and I argued that Anderson's view helps us to better understand how can stereotypes underpin social injustice and how injustice underpinned by stereotypes might be remedied. Finally, drawing on these insights, I developed and defended an account of a form of activism that involves changing the stereotypes that are present in a culture. I called this form of activism stereotype activism.

6. Conclusion: a liberal theory of social justice activism

It is now time to conclude. In this conclusion I will, first, review the argument I have made in this dissertation. Then, I will explain why it is that I call my account a 'liberal theory of social justice activism' and draw out the implications of this theory for the practice of activism in contemporary western societies. Finally, I will say a few words about the value of the contribution I hope to have made in this dissertation and the limitations of that contribution

6.1. Reviewing the argument

The impetus for this project was the observation that, while contemporary political philosophy is replete with rival theories of the perfectly just society, comparatively little attention has been paid to the question of how it is that ideal social justice is to be achieved or advanced towards. A good reason for prioritising ideal over nonideal theory is that one needs to know where one wants to end up in order to assess whether short-term improvements are advancing or impeding one's long-term goal. That said, in the introduction of this dissertation, I argued that the lack of consensus on what perfect social justice looks like should not preclude the development of a relatively ecumenical theory of social justice activism. This is because, starting from where we are, many theories of the perfectly just society can agree on a direction of travel and even on an intermediate destination. I called this intermediate destination 'the core of ideal theory'. The core of ideal theory focused on individuals as the ultimate units of moral concern, it accorded all citizens certain rights, and many of these rights were couched in the language of freedom. The core of ideal theory, as I described it, also expressed a general openness (though certainly not a commitment) to the possibility of inequality in positions when it comes to authority and income, and fair competition for those positions.

When it came to grounding the duty to engage in social justice activism, in chapter 1, I argued that focusing exclusively on any one of the principles that have been advanced as grounds for remedial duties in the literature generates counterintuitive implications which can only be avoided by appealing to multiple principles. Moreover, according to the approach I advocate, and unlike some other accounts, the various values and principles that are invoked to ground remedial duties can combine and interact. This

creates a complicated, pluralistic picture of value and duty. Nevertheless, this pluralistic picture seems to me to account for the complex and messy nature of our moral existence better than a monistic picture – according to which one value explains all duties to engage in social justice activism – would.

In chapter 2, I argued that, in order to understand how to remedy social injustice, we must first understand the different kinds of social phenomena that can underpin injustice. I identified three such phenomena – laws, social norms, and stereotypes – and referred to them as 'the three sites of social justice activism' because I go on to explore the ethics of activism that seeks to intervene in each of them in chapters 3, 4, and 5. The accounts I give of each of these social phenomena are individualistic in the sense that they take these social phenomena to be best described in terms of individual-level properties, such as individuals' beliefs. It is worth noting that it does not follow from the fact that some social phenomena are best described in terms of individual-level properties that the same is true for all social phenomena.

In chapter 3, I explored the ethics of a form of activism that seeks to change the law via means that are either illegal or violent, which I called 'radical activism'. To the extent that the ethics of radical activism has been discussed in the literature, it has typically been considered in the light of the fact that states claim to have political authority. Moreover, often the particular conception of political authority presupposed is one that is binary and grounded in the value of justice. I disaggregated political authority in such a way that revealed that an underappreciated authority relation can exist between a state and its citizens. I called this authority relation 'weak authority', and I grounded it in the value of order and security. More specifically, states are necessary to end the conditions of conflict and insecurity that characterise the state of nature, and a state has weak authority when it has ended these conditions for all individuals living in its territory. I argued that one implication of the presence of weak authority is that the practice of radical activism is more constrained than is standardly supposed.

In chapter 4, I explored the question: how should informal activism be practised? 'Informal activism' involves encouraging people to abandon prevailing unjust social norms and to comply with new, more just social norms instead. My discussion was divided into two parts. First, I critiqued a widely assumed approach to social justice activism, according to which the state is the primary agent of social justice and so citizens' duties to remedy social injustice must be discharged by encouraging the state to change

law and policy. I suggested that the efficacy of this approach is constrained by the fact that certain fundamental interests and the rights they generate create a protected area of social life within which legitimate state action is limited. I called this protected area 'civil society'. Moreover, I argued that this reduced state efficacy means that sometimes informal activism should take place in civil society and need not be mediated by the state.

The second part of the chapter explored how informal activism in civil society should be practised. I suggested that it should be guided by three considerations. First, persuasion is the ideal means of influence in civil society. Second, there is a presumption against what I call 'pressure' – that is, influence that involves seeking to change others' behaviour by physically forcing them to do so or by making it so that their fundamental interests are at risk if they don't. Third, what I call 'price-setting' – which involves seeking to change others' behaviour through the use of incentives where fundamental interests are not at stake – is nonideal but nevertheless permissible.

In chapter 2, I introduced stereotypes, social norms, and law as three mechanisms that can underpin social injustice and, in chapters 3 and 4, I explored how activism that seeks to intervene in law and social norms respectively should be practised. We are very clear how social injustice can be underpinned by law and, though it is less commonly discussed, reasonably clear how injustice can be generated by social norms. However, of the three social phenomena introduced in chapter 2, it seems to me that the process by which social injustice is generated by stereotypes is most subtle and mysterious. Hence, in chapter 5, I sought to better understand the role that stereotypes play in generating social injustice and to draw out the implications of this role for the practice of social justice activism.

My strategy was to begin by examining a recent debate between Elizabeth Anderson and Tommie Shelby regarding racial injustice in the USA and how it should be remedied. Anderson advocates state action to promote racial integration as one of the means by which to remedy this state of affairs but Shelby objects that such an intervention would be illegitimate. I defended Anderson against Shelby's critique by drawing attention to some of the important social science evidence that informs Anderson's view. And I combined the insights that emerged from my analysis of the debate between Anderson and Shelby with further social science evidence to develop an account of a form of activism that I called 'stereotype activism'. Stereotype activism

involves attempting to remedy social injustice by changing the stereotypes that are present in a culture.

6.2. A liberal theory of social justice activism and its implications

Having reviewed the arguments of the preceding chapters, let me now briefly explain why I call my account 'a liberal theory of social justice activism'. Perhaps surprisingly, I will not approach this task by first offering a definition of the word 'liberal', and then highlighting how the theory I have developed in this dissertation meets this definition. This is because the terms 'liberal' and 'liberalism' have been used to refer to many different things – some of which are in tension or inconsistent with each other – and I do not think I am in a position to adjudicate between these different uses to determine which one should be preferred.

Instead, what I will do is highlight certain elements of my account which I think betray a recognisably liberal sensibility. My claim is not that any of these elements is distinctive or exclusive to a liberal outlook, or that together they are held in common by all theories that can be called liberal, or even that, for any theory, possession of this collection of elements is sufficient (though perhaps not necessary) for that theory to be labelled liberal. Rather, my claim is that, *when it comes to a theory of social justice activism in particular*, an account that has these characteristics can appropriately be called liberal.

The characteristics I have in mind are scattered across this dissertation, so let me list them here:

- an assumption that individuals are the ultimate units of moral concern, a focus on rights held by individuals – some of which are couched in terms of freedom, and an openness to inequality and fair competition (see introduction);
- a pluralistic picture of value and duty, and an embracing of the messy and complicated nature of our moral existence (see chapter 1);
- an inclination towards individualistic explanations of social phenomena (see chapter 2);
- a preoccupation with the state as the means by which individuals exit the state nature and so are able to pursue their own ends (see chapter 3);
- an emphasis on the importance of civil society and a defence of persuasion as the ideal means of influence (see chapter 4);

- an approach to solving social problems that is attentive and responsive to social science evidence (see chapter 5).

It is these elements which I want to suggest are, together, sufficient to label my theory of social justice activism liberal. I take this suggestion to be interesting because it seems to me that the liberal outlook is often assumed to be committed to sustaining, or at least to be overly deferential to, the status quo. It would appear to follow from this assumption that, a liberal outlook and a theory of social justice activism would be unhappy bedfellows, because the latter is necessarily concerned with how to transform the status quo. I hope that the theory of social justice activism I have developed in dissertation shows that a liberal outlook need not be committed to defending the status quo. In fact, the liberal perspective characterised by the features just listed contains conceptual and normative resources which enable one to support the need for social justice activism, to ground the duties to engage in social justice activism, to understand some of the different mechanisms that produce social injustice, and to identify the principled and practical constraints that apply to the practice of social justice activism.

Let me try to demonstrate the usefulness of my liberal theory of social justice activism by applying it to something roughly resembling a real-world scenario, and highlighting the implications of such an application. The case I will discuss can only 'roughly resemble' a real-world scenario because the real world is too complex and contains too many relevant variables for any theory to comprehensively and accurately capture. It is also worth reiterating that, especially given the complexity just noted, there is a limit to how much any theory will be able to offer by way of practical guidance, though I will endeavour to offer as much as I can.

Imagine a society that is in various ways very much like many contemporary western societies, but much simpler. Call this society Oceania. The state that governs Oceania has managed to realise order and security in the territory; its citizens no longer pose the threats to one another that create conditions of conflict and insecurity that characterise the state of nature. Nevertheless, Oceania is marred by three, interconnected 'dimensions' of social injustice. First, because Oceania lacks a sufficiently robust social security system, there is a small class of citizens of Oceania who lack the material resources required to live decent, non-degrading lives. This group is also part of a larger class of citizens who do not have fair opportunities to compete for the positions in society

associated with more authority or higher pay. This larger class exists because wealthier citizens of Oceania are able to confer privileges on their children, such as private academic tutors, language lessons, and sports coaches, and the state makes no effort to counteract this activity by providing similar opportunities to children from poorer families or generally investing in the areas where such children live. The result of this is that the children of wealthier citizens have an unfair competitive advantage over the children of poorer citizens. Call this the *economic dimension* of injustice in Oceania.

Second, Oceania contains two genders – men and women – and the citizens of Oceania tend to comply with, endorse, and socialise their children to endorse patriarchal social norms according to which: men should be breadwinners while women must look after the home; sons should be encouraged to pursue higher education, but daughters should not; and the sexual harassment of women by men should be tolerated. These social norms mean that women tend to lack fair opportunities to compete for positions associated with more authority and greater pay, and women's rights to bodily integrity tend to be less secure than men's. Call this the *gender dimension* of injustice in Oceania.

Third, Oceania contains two racial groups – one group of European descent and a second, smaller group of African descent. The recent history of Oceania is marred by serious social injustice in which the citizens of European descent oppressed the citizens of African descent and the state denied citizens of African descent many of their most basic rights and opportunities. The contemporary situation in Oceania is an improvement on the past but the society is still scarred by the legacy of injustice. For one thing, African-Oceanians are overrepresented within the class of citizens who lack the material resources required for a decent life and amongst the larger group who lack fair opportunities to compete for positions that come with more authority and higher pay. One implication of this that European- and African-Oceanians tend to live very segregated lives. Moreover, this legacy of historical injustice has come to inform the stereotypes that European-Oceanians hold about African-Oceanians. Hence, European-Oceanians tend to hold unjust stereotypes which represent African-Oceanians as poor, lazy, unintelligent, incompetent, violent, and prone to crime. One consequence of these stereotypes is that African-Oceanians' rights to bodily integrity are less secure than European-Oceanians. This is because these stereotypes mean that European-Oceanians are more likely to inflict harm on African-Oceanians in cases of mistaken self-defence (see Bolinger 2017). These stereotypes also contribute to the state of affairs in which African-Oceanians lack fair

opportunities. They do so 'directly' by making it so that European-Oceanians are less likely to hire African-Oceanians for jobs. But they also do so 'indirectly' by making it so that European-Oceanians are unlikely to support the kind of policy change that would ensure all had fair opportunities. Call this the *racial dimension* of injustice in Oceania.

The three different dimensions of injustice just described suggest that it is possible to identify three cross-cutting cleavages in Oceania: rich/poor, male/female, and European/African. Because these cleavages are cross-cutting, their various possible combinations create eight different social groups. Each of these social groups stands in a different relationship to the injustice that characterises Oceania, and they interact with the five grounds for remedial duties I discussed in chapter 1 to create a complex picture of the remedial duties held by the citizens of Oceania. Instead of attempting to map these relations and itemise these duties, let me highlight a few especially pertinent categories. And, to make things even simpler, let me assume that all three dimensions of injustice are equally egregious and that, amongst those who contribute to or benefit from injustice, all contribute or benefit to the same degree.

One implication of my intersectional approach to remedial duties is that rich, male, European-Oceanians are likely to have the weightiest duties to engage in social justice activism, while poor, female, African-Oceanians are likely to have the least weighty duties, and other possible social groups stand at various positions in between these two poles. Recall that the weight of an agent's duty refers to the level of cost an agent can be required to bear in order to discharge the duty, and that there is a cap on the level of cost any agent can be required to bear in order to discharge their duty. But recall too that, even if the costs associated with a form of activism exceed the cost cap, the principles that ground the now-extinguished duty will still generate a reason to engage in that form of activism.

Now let me say something about how the various dimensions of the injustice that characterises Oceania should be remedied, starting with the economic dimension. Here, the aim must be to encourage the state to introduce a more robust social security system, and to provide more opportunities for poorer children and invest in the areas where they live. This will involve what I have called formal activism and there are many ways that this might be practised. A low-cost form of formal activism would be to vote for a political party that is in favour of the reforms just described. A moderately costly form of formal activism would be to join a series of demonstrations to support the implementation of such reforms or to make a significant donation to an organisation that advocated for such

reforms. A high cost form of formal activism would be to participate in campaigns of radical activism such as civil disobedience or rioting. Finally, note that, given that the state that governs Oceania has realised order and security in the territory, the citizens of Oceania should avoid acts that risk re-entry into the state of nature. This does not mean that they are never permitted to engage in highly disruptive forms of radical activism. Instead, it means that the citizens of Oceania should exercise more restraint when contemplating such activism than they would need to if they were living in a society where the state had not managed to realise order and security.

Turning now to the gender dimension of injustice in Oceania, we should remember that the capacity of the state to remedy such injustice is curtailed because the injustice is generated by social norms which are incubated and propagated in civil society – that area of social life where the state's ability to intervene is limited. This means that citizens should not just engage in formal activism but should also sometimes engage in informal activism in civil society. There are many possibilities here. A low-cost form of informal activism in civil society might be to begin to use a phrase or slogan developed to critique the unjust social norms in one's everyday life. One – perhaps now dated – example of such a phrase is 'male chauvinist' (see Mansbridge 1999:218-9). A mid-cost form of informal activism in civil society might be for someone to openly criticise unjust social norms in a way that leads to moderately unwelcome consequences for the criticiser. For instance, a man may openly object to and criticise the lecherous and demeaning way his mostly male colleagues talk about and treat the few women in the office, in the knowledge that his male colleagues will ostracise and gossip about him. A high cost form of informal activism might involve forming or joining what I earlier called a 'critical movement'. This may involve living your life somewhat separately from the rest of society and dedicating much of your time and effort to developing a critique of prevailing social norms and a strategy for communicating that critique in a way that is likely to persuade others.

Finally, regarding the racial dimension of injustice in Oceania, here injustice is caused and sustained, in part, by the fact that the legacy of social injustice is embedded in the social environment of Oceania, and that social environment has come to inform the stereotypes about African-Oceanians that European-Oceanians hold. This means that, in order to remedy the racial dimension of injustice in Oceania, it is likely that the stereotypes that are embedded within the social environment of Oceania will need to

change. Typically, only state officials, the leaders of business, educational, or religious institutions, and others who are who in a position to change the character of social environments on a relatively large scale are able to engage in such activities. But stereotype activism is not limited to changing the public representations that are embedded in the social environments. It also covers efforts to change the mental representations that reside in our minds, and this is an activity that ordinary citizens can engage in. Let me focus on this latter kind of stereotype activism.

A low-cost form of such stereotype activism might be to consume media products – such as films, novels, television shows, and videogames – that contain non-stereotypical representations of African-Oceanians. A moderately costly way for an ordinary citizen of Oceania to engage in stereotype activism might be to seek out and join what I will call *temporally limited* social environments which contain non-stereotypical representations of African-Oceanians. Temporally limited social environments are those in which we spend relatively limited amounts of time, such as grocery stores, restaurants, libraries, and gyms. A high-cost form of stereotype activism might be to seek out and join *temporally extended or open-ended* social environments that contain non-stereotypical representations of African-Oceanians. Temporally extended or open-ended social environments are those in which we spend long periods of time, perhaps even indefinitely extended into the future. Examples of temporally extended or open-ended social environments include workplaces, religious associations, and neighbourhoods. I characterise consuming media products and encountering social environments which contain non-stereotypical representations of African-Oceanians in terms of their cost, not because I believe such acts to be inherently costly, but because these activities require investments of time and, at least for some European-Oceanians, may involve enduring social disapproval.

6.3. Contributions and limitations

In this final subsection, I will make some brief remarks about the value of the contributions I hope to have made in this dissertation and the limits of those contributions. As I mentioned earlier, the motivation for this work was the observation that, while there are very many theories of the perfectly just society, comparatively little has been said about how it is that perfect social justice is to be achieved or progressed towards from real-world circumstances that all agree are unjust. The principal

contribution I hope to have made in this dissertation is to have developed an account of the means by which social injustice should be remedied in different circumstances. Moreover, in the hope of offering reflections on social justice activism that are as ecumenical as possible, I have also defended this account of the means by which social injustice should be remedied without first endorsing a fully-developed theory of ideal social justice. Instead, I have only committed to a very modest theory of social justice which I called 'the core of ideal theory'. The central limitations associated with this contribution are that there is only so much practical guidance any general theory will be able to provide, and there will be some who disagree with even the very modest theory of social justice I have assumed, so my account of social justice activism may not be fully ecumenical.

Aside from this principal contribution, there are also a number of subsidiary but standalone contributions which I make in this dissertation. Let me highlight three in particular. The first appears in chapter 2 where I introduce stereotypes, social norms, and law as 'three sites of social justice activism'. There, I also argue for amendments to and an integration of the existing accounts of social norms and law, and so the arguments of chapter 2 also contribute to debates in social ontology. This contribution may be especially significant because the accounts of social norms and law I amend and integrate – Bicchieiri's and Hart's respectively – are two of the leading accounts in the literature. The limitation of this contribution is that, given the broader aims of the dissertation, I was not able to situate it in a wider discussion about the nature of norms where I would be able to show how the integrated account of social norms and law I develop is superior to a recently proposed account that stands as a rival (e.g., Brennan et al. 2012). I hope to pursue this avenue in future work.

The second subsidiary but standalone contribution appears in chapter 3 where I discuss the relationship between political authority and radical activism. There, I invoke the Hohfeldian analysis of rights to argue that the binary account of political authority – the right to rule – presupposed by some political philosophers should be disaggregated. This contribution seems potentially significant because authority is one of the most important concepts in political philosophy and disaggregating it seems likely to yield many interesting implications. However, the limitation is that, while I highlight one additional, underacknowledged authority relation and draw out some of its implications for radical activism, my discussion also implies that there may be further authority

relations, each with their own wider implications, and I was not able to investigate this possibility in this dissertation. Again, I would like to explore this in future work.

The third subsidiary but standalone contribution appears in chapter 4 where I discuss how informal activism should be practised. There, I introduced the idea of civil society: a protected area of social life within which the state's ability to legitimately intervene is restricted. If this protected area is respected as it should be, then the state's ability to engage in informal activism will be limited. Given this, ordinary citizens will sometimes need to engage in informal activism in civil society, and I investigated how such activism should be practised. To emphasise the importance of civil society and to examine the conceptual and normative issues that pertain to relations within this protected space is, I think, to make a potentially significant contribution. This is because there is a tendency within political philosophy to assume that the state-citizen relation is the primary subject of interest. Recognising the status of civil society alerts us to the fact that there are many fascinating and important political issues beyond the state-citizen relation. In fact, some of the most controversial contemporary issues – relating race and gender, say, or 'cancel culture' and free speech – seem to be in large part about how we are to relate each other in civil society. The limitation of this contribution is that, because it appeared within the context of a dissertation that has a very particular focus, it only really scratches the surface of a vast space for enquiry. I hope to explore this topic further in future work.

Let me close by registering a very general contribution I hope to have made in this dissertation. As I have already mentioned, a great deal of attention in political philosophy is focused on describing the perfectly just society that should be our end, but much less attention is paid to the question of the means by which we should reach that end. This dissertation is an attempt to contribute to redressing that imbalance, but I am sure that there is much, much more to say. I hope that this dissertation goes some way to spurring others to ensure that political philosophy is able answer questions that pertain to means as well as ends.

Bibliography

Anderson, Elizabeth. 2009. 'Towards a Non-Ideal, Relational Methodology for Political Philosophy: Comments on Schwartzman's "Challenging Liberalism",' *Hypatia*, 24(4): 130-145.

Anderson, Elizabeth. 2010. *The Imperative of Integration*. Woodstock: Princeton University Press.

Anderson, Elizabeth. 2014. 'Social Movements, Experiments in Living, and Moral Progress: Case Studies from Britain's Abolition of Slavery,' *The Lindley Lecture*. Available online: https://kuscholarworks.ku.edu/handle/1808/14787 (Last accessed: 4th August 2021).

Anderson, Elizabeth. 2018. 'Review of *Dark Ghettos: Injustice, Dissent, and Reform* by Tommie Shelby,' *Mind*, 127(505): 276-284.

Appiah, Kwame Anthony. 2010. *The Honor Code: How Moral Revolutions Happen*. London: W. W. Norton & Company.

Arendt, Hannah.1970. *On Violence*. London: Allen Lane The Penguin Press.

Barnhill, Anne. 2014. 'What is Manipulation?,' in Christian Coons and Michael Weber, eds., *Manipulation: Theory and Practice*, 51-72. Oxford University Press.

BBC Trending. 2015. 'Damonsplaining: Matt Damon accused of insensitivity,' *BBC News*, 16th September. Available online: https://www.bbc.co.uk/news/blogs-trending-34251224 (Last accessed: 4th August 2021).

Bertrand, Marianne and Mullainathan, Sendhil. 2004. 'Are Emily and Gregg More Employable than Lakisha and Jamal? A Field Experiment on Labour Market Discrimination,' *American Economic Review*, 94(4): 991-1013.

Bicchieri, Cristina. 2006. *The Grammar of Society: The Nature and Dynamics of Social Norms*. Cambridge: Cambridge University Press.

Bicchieri, Cristina. 2017. *Norms in the Wild: How to Diagnose, Measure, and Change Social Norms*. Oxford: Oxford University Press.

Bicchieri, Cristina and Marini, Annalisa. 2015. *Female Genital Cutting: Fundamentals, Social Expectations and Change*. Available online: https://mpra.ub.uni-muenchen.de/72927/ (Last accessed: 4th August 2021).

Bicchieri, Cristina et al. 2018. 'Social Norms,' in Edward N. Zalta, ed., *The Stanford Encyclopedia of Philosophy*. Available online: https://plato.stanford.edu/entries/social-norms/ (Last accessed: 4th August 2021).

Bolinger, Renée Jorgensen. 2017. 'Reasonable Mistakes and Regulative Norms: Racial Bias in Defensive Harm,' *The Journal of Political Philosophy*, 25(2): 196-217.

Bray, Mark. 2017. 'Forum Response: Notes for Antifa from a Former "Terrorist",' *Boston Review*, 29th November. Available online: http://bostonreview.net/forum/bray-notes-etzioni-antifa-historian (Last accessed: 4th August 2021).

Brennan, Geoffrey et al. 2012. *Explaining Norms*. Oxford: Oxford University Press.

Brook, Erynn. 2018. 'Is the term "mansplaining" sexist?' You asked Google – here's the answer,' *The Guardian*, 6th June. Available online: https://www.theguardian.com/commentisfree/2018/jun/06/is-the-term-mansplaining-sexist-google-autocomplete (Last accessed: 4th August 2021).

Brownlee, Kimberley. 2012. *Conscience and Conviction: The Case for Civil* Disobedience. Oxford: Oxford University Press.

Brownlee, Kimberley. 2013. 'A Human Right Against Social Deprivation,' *The Philosophical Quarterly*, 63(251): 199-222.

Brownlee, Kimberley. 2015. 'Freedom of Association: It's Not What You Think,' *Oxford Journal of Legal Studies*, 35(2): 267-282.

Brownlee, Kimberley. 2016. 'Ethical Dilemmas of Sociability,' *Utilitas*, 28(1): 54-72.

Brownstein, Michael. 2019. 'Implicit Bias,' in Edward N. Zalta, ed., *The Stanford Encyclopedia of Philosophy*. Available online: https://plato.stanford.edu/entries/implicit-bias/ (Last accessed: 4th August 2021).

Buchanan, Allen. 2002. 'Political Legitimacy and Democracy,' *Ethics*, 112(4): 689-719.

Buchanan, Allen. 2013. 'The Ethics of Revolution and Its Implications for the Ethics of Intervention,' *Philosophy and Public Affairs*, 41(4): 292-323.

Buehler, James W. 2017. 'Racial/Ethnic Disparities in the Use of Lethal Force by US Police, 2010-2014,' *American Journal of Public Health*, 107(2): 295-297.

Butt, Daniel. 2007. 'On Benefitting from Injustice,' *Canadian Journal of Philosophy*, 37(1): 129-152.

Caney, Simon. 2010. 'Climate change and the duties of the advantaged,' *Critical Review of International Social and Political Philosophy*, 13(1): 203-208.

Chambers, Simone and Kopstein, Jeffrey. 2006. 'Civil Society and the State,' in John S. Dryzek et al., eds., *The Oxford Handbook of Political Theory*, 363-381. Oxford: Oxford University Press.

Chenoweth, Erica and Stephan, Maria J. 2011. *Why civil resistance works: the strategic logic of nonviolent conflict*. Chichester: Columbia University Press.

Christiano, Tom. 2012. 'Authority,' in Edward N. Zalta, ed., *The Stanford Encyclopedia of Philosophy*. Available online: https://plato.stanford.edu/entries/authority/ (Last accessed: 4th August 2021).

Cogan, Marin. 2012. 'The Mittsplainer: An Alternative Theory of Mitt Romney's Gaffes,' *GQ*, 1st August. Available online: https://www.gq.com/story/the-mittsplainer-an-alternate-theory-of-mitt-romneys-gaffes (Last accessed: 4th August 2021).

Cohen, G. A. 1997. 'Where the Action is: On the Site of Distributive Justice,' *Philosophy & Public Affairs*, 26(1): 3-20.

Cohen, G. A. 2009. *Why Not Socialism?*. Oxford: Princeton University Press.

Coons, Christian and Weber, Michael. 2014. 'Manipulation: Investigating the Core Concept and Its Moral Status,' in Christian Coons and Michael Weber, eds., *Manipulation: Theory and Practice*, 1-16. Oxford University Press.

Correll, Joshua et al. 2007. 'The influence of stereotypes on decisions to shoot,' *European Journal of Social Psychology*, 37: 1102-1117.

Crenshaw, Kimberlé. 1989. 'Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics,' *University of Chicago Legal Forum*, 1: 139-167.

Dagger, Richard and Lefkowitz, David. 2021. 'Political Obligation,' in Edward N. Zalta, ed., *The Stanford Encyclopedia of Philosophy*. Available online: https://plato.stanford.edu/entries/political-obligation/ (Last accessed: 4th August 2021).

Dasgupta, Nilanjana and Asgari, Shaki. 2004. 'Seeing is believing: Exposure to countersterotypic women leaders and its effect of the malleability of automatic gender stereotyping,' *Journal of Experimental Social Psychology*, 40: 642-658.

Dasgupta, Nilanjana and Greenwald, Anthony G. 2001. 'On the Malleability of Automatic Attitudes: Combating Automatic Prejudice With Images of Admired and Disliked Individuals,' *Journal of Personality and Social Psychology*, 81(5): 800-814.

de Sousa, Ronald. 1987. *The Rationality of Emotion*. London: The MIT Press.

Delmas, Candice. 2014. 'Political Resistance: A Matter of Fairness,' *Law and Philosophy*, 33: 465-488.

Delmas, Candice. 2018. *A Duty to Resist: When Disobedience Should be Uncivil*. Oxford: Oxford University Press.

Dixon, Travis L and Linz, Daniel. 2000. 'Overrepresentation and Underrepresentation of African Americans and Latinos as Lawbreakers on Television News,' *Journal of Communication*, 50(2): 131-154.

Edwards, Griffin and Rushin, Stephen. 2018. *The Effect of President Trump's Election on Hate Crimes*. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3102652 (Last accessed: 4th August 2021).

Eagly, Alice H. and Karau, Steven J. 2002. 'Role Congruity Theory of Prejudice Towards Female Leaders,' *Psychological Review*, 109(3): 573-598.

Finlay, Christopher J. 2010. 'Legitimacy and Non-state Political Violence,' *The Journal of Political Philosophy*, 18(3): 287-312.

Fiske, Susan et al. 1991. 'Social Science Research on Trial: Use of Sex Stereotyping Research in Price Waterhouse v. Hopkins,' *American Psychologist*, 46(10): 1049-1060.

Galligan, Denis J. 2003. 'Legal Failure: Law and Social Norms in Post-Communist Europe,' in Denis J. Galligan and Marina Kurkchiyan, eds., *Law and Informal Practices*, 1-23. Oxford: Oxford University Press.

Gaus, Gerald. 2011. *The Order of Public Reason: A Theory of Freedom and Morality in a Diverse and Morally Bounded World*. Cambridge: Cambridge University Press.

Ghandnoosh, Nazgol. 2014. *Race and Punishment: Racial Perceptions of Crime and Support for Punitive Policies*. Available online: https://www.sentencingproject.org/publications/race-and-punishment-racial-perceptions-of-crime-and-support-for-punitive-policies/ (Last accessed: 4th August 2021).

Gilabert, Pablo. 2017. 'Justice and Feasibility,' in Michael Weber and Kevin Vallier, eds., *Political Utopias: Contemporary Debates*, 95-124. Oxford: Oxford University Press.

Gilens, Martin. 1996. 'Race and Poverty in America: Public Misperceptions and the American News Media,' *The Public Opinion Quarterly*, 60(4): 515-541.

Google Trends. 2021. *Explore* (Search term: 'mansplain'). Available online: https://trends.google.com/trends/explore?date=all&q=mansplain (last accessed 4th August 2021).

Greenwald, Anthony G et al. 1998. 'Measuring Individual Differences in Implicit Cognition,' *Journal of Personality and Social Psychology*, 74(6): 1464-1480.

Griffin, James. 1988. *Well-Being: Its Meaning, Measurement, and Moral Importance*. Oxford: Oxford University Press.

Hanna, Jason. 2015. 'Libertarian Paternalism, Manipulation, and the Shaping of Preferences,' *Social Theory and Practice*, 41(4): 618-643.

Haque, Adil Ahmad. 2017. *Law and Morality at War*. Oxford: Oxford University Press.

Hart, H. L. A. 1982. *Essays on Bentham: Jurisprudence and Political Theory*. Oxford: Oxford University Press.

Hart, H. L. A. 2012. *The Concept of Law* (Third Edition). Oxford: Oxford University Press.

Hayward, Clarissa Rile. 2017. 'Responsibility and Ignorance: On Dismantling Structural Injustice,' *The Journal of Politics*, 79(2): 396-408.

Hayward, Clarissa Rile. 2020. 'Disruption: What Is It Good For?,' *The Journal of Politics*, 82(2): 448-459.

Heilman, Madeline et al. 2004. 'Penalties for Success: Reaction to Women Who Succeed at Male Gender-Typed Tasks,' *Journal of Applied Psychology*, 89(3): 416-427.

Hohfeld, Wesley Newcomb. 1923. *Fundamental Legal Conceptions as Applied in Judicial Reasoning and Other Legal Essays*. London: Oxford University Press.

Horton, John. 2006. 'In Defence of Associative Political Obligations: Part One,' *Political Studies*, 54(3): 427-443.

Horton, John. 2007. 'In Defence of Associative Political Obligations: Part Two,' Political Studies, 55(1): 1-19.

House of Commons. 2013. *Parliamentary Debates (Hansard)*, 8th January. Available online: https://publications.parliament.uk/pa/cm201213/cmhansrd/chan92.pdf (Last accessed: 4th August 2021).

Huemer, Michael. 2013. *The Problem of Political Authority: An Examination of the Right to Coerce and the Duty to Obey*. London: Palgrave Macmillan.

Klosko, George. 2005. *Political Obligations*. Oxford: Oxford University Press.

Kahan, Dan M. 2000. 'Gentle Nudges vs. Hard Shoves: Solving the Sticky Norms Problem,' *The University of Chicago Law Journal*, 67(3): 607-645.

King Jr., Martin Luther. 1963. *Letter From a Birmingham Jail*. Available online: http://okra.stanford.edu/transcription/document_images/undecided/630416-019.pdf (Last accessed: 4th August 2021).

King Jr., Martin Luther. 1964. *The Quest for Peace and Justice* (Nobel Lecture, delivered on 11th December). Available online: http://www.nobelprize.org/nobel_prizes/peace/laureates/1964/king-lecture.html (Last accessed: 4th August 2021).

Kuran, Timur. 1995. *Private Truths, Public Lies: The Social Consequences of Preference Falsification*. London: Harvard University Press.

Kurkchiyan, Marina. 2003. 'The Illegitimacy of Law in Post-Soviet Societies,' in Denis J. Galligan and Marina Kurkchiyan, eds., *Law and Informal Practices*, 25-46. Oxford: Oxford University Press.

Ladenson, Robert. 1980. 'In Defence of a Hobbesian Conception of Law,' *Philosophy & Public Affairs*, 9(2): 134-159.

Lamond, Grant. 2000. 'The Coerciveness of Law,' *Oxford Journal of Legal Studies*, 20(1): 39-62.

Laurence, Ben. 2020. 'The Question of the Agent of Change,' *The Journal of Political Philosophy*, 28(4): 355-377.

Lomasky, Loren E. 1987. *Persons, Rights, and the Moral Community*. Oxford: Oxford University Press.

Lyons, David. 2015. 'Natural Duty of Justice,' in Jon Mandle and David A. Reidy, eds., *The Cambridge Rawls Lexicon*, 551-552. Cambridge: Cambridge University Press.

Mackie, Gerry. 1996. 'Ending Footbinding and Infibulation: A Conventional Account,' *American Sociological Review*, 61(6): 999-1017.

Mackie, Gerry. 2017. 'Effective Rule of Law Requires Construction of a Social Norm of Legal Obedience,' in Carlo Tognato, ed., *Cultural Agents Reloaded: The Legacy of Antanas Mockus*, 313-334. Cambridge: Harvard University Press.

Mackie, Gerry and LeJaune, John. 2009. 'Social dynamics of abandonment of harmful practices: a new look at the theory,' *UNICEF Innocenti Research Centre Working Paper*. Available online: https://www.unicef-irc.org/publications/558-social-dynamics-of-abandonment-of-harmful-practices-a-new-look-at-the-theory.html (Last accessed: 4th August 2021).

Madva, Alex. 2020. 'Integration, Community, and the Medical Model of Social Injustice,' *Journal of Applied Philosophy*, 37(2): 211-232.

Mandela, Nelson. 1964. *"I am Prepared to Die" – Statement at the opening of the defence case in the Rivonia Trial* (delivered on 20th April). Available online: http://db.nelsonmandela.org/speeches/pub_view.asp?pg=item&ItemID=NMS010&txtstr=prepared%20to%20die (Last accessed: 4th August 2021).

Mansbridge, Jane. 1996. 'Using Power/Fighting Power: The Polity,' in Selya Benhabib, ed., *Democracy and Difference: Contesting the Boundaries of the Political*, 46-66. Chichester: Princeton University Press.

Mansbridge, Jane. 1999. 'Everyday Talk in the Deliberative System,' in Stephen Macedo, ed., *Deliberative Politics: Essays on Democracy and Disagreement*, 211-239. Oxford: Oxford University Press.

Marmor, Andrei. 2015. 'What is the Right to Privacy?,' *Philosophy & Public Affairs*, 43(1): 3-26.

McGarty, Craig et al. 2002. 'Social, cultural and cognitive factors in stereotype formation,' in Craig McGarty et al., eds., *Stereotypes as Explanations: The formation of meaningful beliefs about social groups*, 1-15. Cambridge: Cambridge University Press.

McMahan, Jeff. 2016. 'Preventive War and the Killing of the Innocent,' in Richard Sorabji and David Rodin, eds., *The Ethics of War: Shared Problems in Different Traditions*, 169-190. London: Routledge.

Meckled-Garcia, Saladin. 2008. 'On the Very Idea of Cosmopolitan Justice: Constructivism and International Agency,' *The Journal of Political Philosophy*, 16(3): 245-271.

Miller, David. 2001. 'Distributing Responsibilities,' *The Journal of Political Philosophy*, 9(4): 453-471.

Miller, David. 2005. 'Reasonable Partiality Towards Compatriots,' *Ethical Theory and Moral Practice*, 8(1/2): 63-81.

Miller, David. 2007. *National Responsibility and Global Justice*. Oxford: Oxford University Press.

Morreall, John. 1991. 'The Justifiability of Violent Civil Disobedience,' in Hugo Adam Bedau, ed., *Civil Disobedience in Focus*, 130-143. London: Routledge.

Nagel, Thomas. 1998. 'Concealment and Exposure,' *Philosophy & Public Affairs*, 27(1): 3-30.

Noggle, Robert. 1996. 'Manipulative Actions: A Conceptual and Moral Analysis,' *American Philosophical Quarterly*, 33(1): 43-55.

Noggle, Robert. 2018. 'Manipulation, salience, and nudges,' *Bioethics*, 32: 164-170.

OED. 2021. *Violence*. Available online: http://www.oed.com/view/Entry/223638?isAdvanced=false&result=1&rskey=CfYLQ0& (Last accessed: 4th August 2021).

O'Neill, Onora. 1998. 'Rights, Obligations, and Needs,' in Gillian Brock, ed., *Necessary Goods: Our Responsibilities to Meet Others' Needs*, 95-112. Oxford: Rowman & Littlefield Publishers.

O'Neill, Onora. 2001. 'Agents of Justice,' *Metaphilosophy*, 32(1/2): 180-195.

The Opportunity Agenda. 2011. *Media Representations and Impact on the Lives of Black Men and Boys*. Available online: https://www.opportunityagenda.org/explore/resources-publications/social-science-literature-review (Last accessed: 4th August 2021).

Pasternak, Avia. 2010. 'Sharing the costs of political injustices,' *Politics, Philosophy, and Economics*, 10(2): 188-210.

Pasternak, Avia. 2019. 'Political Rioting: A Moral Assessment,' *Philosophy & Public Affairs*, 46(4): 384-418.

Payne, B. Keith et al. 2017. 'The Bias of Crowds: How Implicit Bias Bridges Personal and Systemic Prejudice,' *Psychological Inquiry*, 28(4): 233-248.

Poama, Andrei. 2020. 'Staying in or moving out? Justice and the abolition of the dark ghetto,' *European Journal of Political Theory*, 19(1): 128-138.

Rachels, James. 1975. 'Why Privacy is Important,' *Philosophy & Public Affairs*, 4(4): 323-333.

Rawls, John. 1999a. *A Theory of Justice* (Revised Edition). Oxford: Oxford University Press.

Rawls, John. 1999b. *The Law of Peoples: with "The Idea of Public Reason Revisited"*. London: Harvard University Press.

Rawls, John. 2005. *Political Liberalism* (Expanded Edition). Chichester: Columbia University Press.

Raz, Joseph. 1986. *The Morality of Freedom*. Oxford: Oxford University Press.

Renzo, Massimo. 2011. 'State Legitimacy and Self Defence,' *Law and Philosophy*, 30: 575-601.

Renzo, Massimo. 2012. 'Associative Responsibilities and Political Obligation,' *The Philosophical Quarterly*, 62(246): 106-127.

Robeyns, Ingrid. 2012. 'Are transcendental theories of justice redundant?,' *Journal of Economic Methodology*, 19(2): 159-163.

Rochon, Thomas R. 1998. *Culture Moves: Ideas, Activism, and Changing Values*. Chichester: Princeton University Press.

Ronzoni, Miriam. 2007. 'Two Concepts of The Basic Structure, And Their Relevance to Global Justice,' *Global Justice: Theory Practice Rhetoric*, 1: 68-85.

Ronzoni, Miriam. 2008. 'What Makes the Basic Structure Just?,' *Res Publica*, 14: 203-218.

Rooth, Dan-Olof. 2010. 'Automatic associations and discrimination in hiring: Real world evidence,' *Labour Economics*, 17: 523-534.

Rudman, Laurie A. and Glick, Peter. 2001. 'Prescriptive Gender Stereotypes and Backlash Towards Agentic Women,' *Journal of Social Issues*, 57(4): 743-762.

Ryerson, James. 2016. 'Race in America After the Great Migration,' *The New York Times*, 18th November. Available online: https://www.nytimes.com/2016/11/20/books/review/race-in-america-after-the-great-migration.html (Last accessed: 4th August 2021).

Sangiovanni, Andrea. 2015. 'Solidarity as Joint Action,' *Journal of Applied Philosophy*, 32(4): 340-359.

Sarkin, Jeremy and Koenig, Mark. 2010. 'Ending Caste Discrimination in India: Human Rights and the Responsibility to Protect (R2R) Individuals and Groups from Discrimination at the Domestic and International Levels,' *The George Washington International Law Review*, 41(3): 541-576.

Sawyer, Wendy and Wagner, Peter. 2020. *Mass Incarceration: The Whole Pie 2020*. Available online: https://www.prisonpolicy.org/factsheets/pie2020_allimages.pdf (Last accessed: 4th August 2021).

Schouten, Gina. 2013. 'Restricting Justice: Political Interventions in the Home and in the Market,' *Philosophy & Public Affairs*, 41(4): 357-388.

Sen, Amartya. 2006. 'What Do We Want from a Theory of Justice?,' *The Journal of Philosophy*, 103(5): 215-238.

Shelby, Tommie. 2016. *Dark Ghettos: Injustice, Dissent, and Reform*. London: Belknap Press of Harvard University Press.

Sifton, Sam and Barrett, Grant. 2010. 'The Words of the Year,' *The New York Times*, 18th December. Available online: https://www.nytimes.com/2010/12/19/weekinreview/19sifton.html (Last accessed: 4th August 2021).

Silvermint, Daniel. 2013. 'Resistance and Well-being,' *The Journal of Political Philosophy*, 21(4): 405-425.

Simmons, A. John. 1979. *Moral Principles and Political Obligations*. Chichester: Princeton University Press.

Simmons, A. John. 2001. *Justification and Legitimacy: Essays on Rights and Obligations*. Cambridge: Cambridge University Press.

Simmons, A. John. 2010. 'Ideal and Nonideal Theory,' *Philosophy & Public Affairs*, 38(1): 5-36.

Singer, Peter. 1972. 'Famine, Affluence, and Morality,' *Philosophy and Public Affairs*, 1(3): 229-243.

Solnit, Rebecca. 2008. 'Men Explain Things to Me; Facts Didn't get in Their Way,' *Common Dreams*, 13th April. Available online: https://www.commondreams.org/views/2008/04/13/men-explain-things-me-facts-didnt-get-their-way (Last accessed: 4th August 2021).

Sperber, Dan. 1996. *Explaining Culture: A Naturalistic Approach*. Oxford: Blackwell Publishing.

Spiekermann, Kai. 2015. 'Review of *Explaining Norms* by Brennan et al.,' *Economics and Philosophy*, 31(1): 174-181.

Sreenivasan, Gopal. 2007. 'Health and justice in our non-ideal world,' Politics, Philosophy & Economics, 6(2): 218-236.

Stemplowska, Zofia and Swift, Adam. 2012. 'Ideal and Nonideal Theory,' in David Estlund ed., *The Oxford Handbook of Political Philosophy*, 373-390. Oxford: Oxford University Press.

Sunstein, Cass R. 1996. 'Social Norms and Social Roles,' *Columbia Law Review*, 96(4): 903-968.

Swift, Adam. 2008. 'The Value of Philosophy in Nonideal Circumstances,' *Social Theory and Practice*, 34(3): 363-387.

Thompson, Judith Jarvis. 1990. *The Realm of Rights*. London: Harvard University Press.

Walzer, Michael. 1970. *Obligations: Essays on Disobedience, War, and Citizenship*. London: Harvard University Press.

Weber, Max. 1994[1919]. 'The Profession and Vocation of Politics,' in *Political Writings*. Cambridge: Cambridge University Press.

Wellman, Christopher. 1996. 'Liberalism, Samaritanism, and Political Legitimacy,' *Philosophy & Public Affairs*, 25(3): 211-237.

Wittenbrink, Bernd et al. 2001. 'Spontaneous Prejudice in Context: Variability in Automatically Activated Attitudes,' *Journal of Personality and Social Psychology*, 81(5): 815-827.

Wolff, Jonathan. 1998. 'Fairness, Respect, and the Egalitarian Ethos,' *Philosophy & Public Affairs*, 27(2): 97-122.

Wolff, Jonathan and de-Shalit, Avner. 2007. *Disadvantage*. Oxford: Oxford University Press.

Wood, Martin et al. 2009. 'A test for racial discrimination recruitment practice in British cities,' *Department for Work and Pensions Research Report No 607*. Available online: http://www.natcen.ac.uk/media/20541/test-for-racial-discrimination.pdf (Last accessed: 4th August 2021).

Young, Iris Marion. 2011. *Responsibility for Justice*. Oxford: Oxford University Press.