What Does Justice Require of Me?

The Individual Demands of Social Justice, and the Paradox of Collective Harm

Rowan Mellor

UCL

PhD

I, Rowan Mellor, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

For Annie and Des

Already the Great Kahn was leafing through his atlas, over the maps of the cities that menace in nightmares and maledictions: Enoch, Babylon, Yahooland, Batua, Brave New World.
He said: "It is all useless, if the last landing place can only be the infernal city, and it is there that, in ever-narrowing circles, the current is drawing us."
And Polo said: "The inferno of the living is not something that will be; if there is one, it is what is already here, the inferno where we live every day, that we form by being together. There are two ways to escape suffering it. The first is easy for many: accept the inferno and become such a part of it that you can no longer see it. The second is risky and demands constant vigilance and apprehension: seek and learn to recognize who and what, in the midst of the inferno, are not inferno, then make them endure, give them space."

Italo Calvino, *Invisible Cities*

Acknowledgements

This thesis is not the most important thing in my life; and I would like to thank first the friends and family who have helped to remind me of this fact. This past year has been extraordinary in many ways, not least because I have spent most of it with one person. Very special thanks (and apologies) are owed, therefore, to Becky.

I have now been a member of UCL's Department of Philosophy for a decade (first as an undergraduate student, then as graduate, and now as a teacher), and for most of this time I have been thinking about topics related to those discussed in this thesis. It is therefore difficult to trace all of the individuals who, in one way or another, have had an influence on it. Instead, I will simply mention some of the people whose contributions stand out. First and foremost, I want to register a special debt of gratitude to Véronique Munoz-Dardé. Véronique's contribution to what follows goes far beyond that of merely highlighting errors, suggesting argumentative moves, recommending readings etc. (all of which she has done). She has instilled in me a sense of philosophy as a practice: how it is done, and how it is valuable. If it were not for her insight and guidance, this thesis would not exist. I am also indebted to the following people (and probably many more), numerous conversations with whom have helped to give shape to the thoughts expressed here: Han van Wietmarschen, Joe Horton, Ulrike Heuer, Nikhil Venkatesh, Jessica Fischer, Hannah Carnegy-Arbuthnott, Showkat Ali, and Dan Guillery. Finally, I would like to thank everyone who, over the years, has attended the Philosophy Work in Progress seminars, and the meetings variously titled (amongst other things) 'MopWip', 'Moral and Political Workshop', and 'Véronique's Babies'. I have presented work at these sessions on more occasions than I can remember, and have always come away with more insights than I can do justice to.

Abstract

For many of us, the existence of social injustice is a source of disquiet. Recognising that social conditions are unjust seems to have implications for how each of us ought to act. Yet precisely what those implications are is difficult to establish. Much social injustice can be thought of as 'structural', where this implies that it is the cumulative result of largescale behavioural patterns in which a given individual's actions appear negligible. As such, structural injustice seems to be something for which no one can be held personally responsible. In this thesis, I provide accounts of both the personal requirements of social justice and the wrong of collective harm. The result is a picture of our responsibilities with respect to justice which specifically addresses conditions of structural injustice. Part 1 considers the nature of structural injustice, and delineates the personal requirements implied by particular institutional principles. I argue that the existence of structural injustice supports the view that our principles of social justice should apply primarily at the institutional level, but that this does not release individuals from responsibility. Part 2 tackles the question of whether we always have a reason to refrain from collectively harmful actions. It might seem plausible to answer in the affirmative. However, I point out that this conflicts with another seemingly plausible thought: that it is possible for a series of actions to be both collectively harmful and severally harmless. I offer a novel solution to this problem. I argue that lawmakers can be obligated to regulate collectively harmful behaviour, even if individuals have no pre-legal reason to avoid it. As such, whilst we might not always have a reason to refrain from collectively harmful actions, we can nonetheless have a moral claim on institutional actors to protect us from their joint effects.

Impact Statement

The existence of social injustice faces us, as individual members of society, with urgent questions about what we ought to do. The fact that social conditions are unjust seems to imply that at least someone ought to do something about it. But *who*, and *what*? In this thesis, I explore the moral requirements which individuals bear in the face of social injustice. Of particular concern are conditions of injustice which can be described as 'structural', where this implies that those conditions are the cumulative result of largescale behavioural patterns. I offer accounts of both the individual requirements of social justice and the wrong of collective harm in order to develop a view of our responsibilities with respect to structural injustice.

The thesis makes novel contributions to several areas of moral and political philosophy. Part 1 participates in debates surrounding the subject of the principles of social justice, the interaction between principles of equal opportunity and parental partiality, and the nature of structural injustice. The principal academic contribution is a new account of the theoretical significance of the concept of structural injustice, and the moral requirements which it implies. Part 2 provides an original theory of the wrong of collective harm. Most contributors to this debate argue that it is morally wrong for an individual to perform one of several collectively harmful actions, even if that action causes no harm considered singly. By contrast, I argue that we should reject this claim. I agree with other theorists that we have a valid moral claim against being harmed as the cumulative result of others' actions. However, on my view, this does not stem from the fact that people are generally obligated to avoid acting as such. Rather, it stems from the fact that legislators are obligated to regulate collectively harmful behaviour, even if individuals have no pre-legal reason to avoid it.

The views developed in this thesis also have implications for issues outside of academic philosophy. Principal among these are urgent areas of public policy, such as the climate crisis. The harms caused by anthropogenic climate

change can be viewed as instances of collective harm and structural injustice: they are the cumulative result of many individuals' uncoordinated actions. As such, the accounts offered here provide the theoretical groundings for an ethical code to guide policymakers in tackling these problems. The thesis also addresses a pressing concern of the current political climate. In the months following the murder of George Floyd by police officer Derek Chauvin in May 2020, it was difficult to find a (reputable) newspaper or cultural magazine which did not include some mention of 'structural racism' or 'structural injustice'. Precisely how to theorise these notions, and what they imply for how we, as individuals, ought to act, is one of the core questions of this thesis.

Contents

Introduction

Between 1979 and 1980, the photographer David Goldblatt took a series of portraits in the South African city of Boksburg. Compiling a retrospective of his work years later, he looked back on the experience: "White and black: locked into a system of manic control and profound immorality. To draw breath there was to be complicit" (Goldblatt 2018, p. 197).

Goldblatt's retrospect captures something of the subject of this thesis. Here is a platitude of ethical thought: you ought not to do wrong to others. Here is another: absent countervailing considerations, you ought not to cause harm to others. And here is something short of a platitude, but still a commonplace: if you wrong someone else, or cause them harm, then you ought to do something to make it up to them. Each of these three claims describes a way for a person to be connected to a wrong or harm: one can 'do' a wrong, or 'cause' a harm. Furthermore, each of them also describes a moral requirement associated with that way of being connected to a wrong/harm: we are required not to do wrong to others and, other things being equal, not to cause harm to them; and if we transgress these requirements, then we are required to make amends with whomever we have wronged or harmed.

So far, so simple. However, as Goldblatt intimates, there seem to be a variety of further, less direct ways in which an individual can be connected to a wrong or harm. What, if any, moral requirements are associated with such connections?

The central idea which underlies this thesis (the 'thesis of my thesis', as it's sometimes called at UCL) can be understood as follows. We can gain a clearer understanding of the moral requirements associated with certain indirect ways of being connected to a wrong or harm by appreciating the ways in which the duties[1] we owe to one another can depend upon the social institutions in which

[1] Throughout this thesis, I will use the terms 'duty', 'obligation', and 'moral requirement' interchangeably. For a possible way of differentiating the notions of duty and obligation, see (Simmons 1979, chap. 1).

we are embedded. We find ourselves encased in social institutions: more or less formal systems of rules which regulate our behaviour, define the terms of certain forms of interaction, and determine how the goods of social cooperation get distributed. This thought is not new; nor is the idea that it importantly affects how we should understand the duties which we bear to each other. But there is a thought in the vicinity the implications of which, at least in certain contemporary debates, have not been fully grasped: the institutions we inhabit can determine the moral claims which we can legitimately make of others. By a moral claim, I mean some interest of a person which implies a corresponding duty on others. For instance, if I have a claim against being treated in some way, then others have a duty not to treat me in that way; or if I have a claim to some good, then others at least have a duty not to deny me it. The thought at the core of this thesis is that the claims we can make of one another, and thus the duties we are subject to, depend on the social institutions in which we are embedded, and how they are organised; and moreover, that recognising this can help us to make sense of what is required of us in cases of indirect connection with a wrong or harm.

This is best demonstrated by reviewing the contents of this thesis in a non-linear fashion; let's start with Part 2. Here, my explicit question is: Does the fact that an action would be one of several which will collectively cause harm to others constitute, or imply, a reason not to perform it? This is a version of my broader, underlying question: performing one among several collectively harmful actions is a way of being causally connected to that harm without being *the* one who causes it; and having a reason not to do something is, I assume, a precondition of being morally required not to do it. It might seem quite natural to suppose that we do have a reason not to engage in collectively harmful behaviour. However, I argue in Chapter 4 that this plausible-sounding thought is threatened when we consider that several actions might together cause a harmful outcome, without any one of them making any difference to it. The possibility of such cases generates friction between two intuitive moral judgements. On the one hand, we are inclined to say that, other things being equal, we ought not to do things which together harm others. But, on the other, we also find it difficult to explain why a given individual

ought to refrain from an action if, on its own, it will make no difference to the harm in question.

How should we resolve this tension? Chapter 4 considers and rejects some of the foremost solutions in the literature. In Chapter 5, I suggest that we try a fresh approach. I argue that we can make moral claims on lawmakers to regulate certain forms of behaviour, even if, but for such regulation, we could not make claims on others not to behave in those ways. Correspondingly, lawmakers can be obligated to regulate some act, even if the individuals whose activity they ought to regulate have no pre-legal reason to avoid acting as such. This is an example of how the claims which we can make on, and the duties we bear to, one another can depend on the social institutions in which we are embedded. A legal system is a social institution; it comprises of a system of rules which not only regulate the actions of those within its jurisdiction, but also define who may enact changes to those first-order rules, and how such changes may be enacted.[2] As such, if we can make claims on legislators to regulate some act, but not on others to refrain from actions of that type when they go unregulated, then whether or not such an action flouts a legitimate moral claim may depend on the social institutions in place: specifically, on whether there is a legal system, and how it is organised.

This point makes possible a solution to the problem set out in Chapter 4. I argue that we should reject the claim that we always have a reason to refrain from an action if it would be one of several which will collectively harm others. However, we need not therefore deny that we have a moral claim against being harmed as the joint result of others' actions. Lawmakers, I argue, can be obligated to regulate collectively harmful behaviour, even if individuals have no pre-legal reason to refrain from it. As such, given that we possess a legal system, people can have a legitimate moral objection to collectively caused harm, even if, but for the law, they would lack any reason to avoid collectively harmful actions.

[2] See (Hart 2012, chap. 5)

Attentive readers will have noticed that I have not yet mentioned the question which I have chosen as my title: What does justice require of me? More specifically: What does the ideal of social justice require of the individual members of a society? Here is one way of understanding how it relates to the foregoing discussion. Social injustice is a wrong: those who are subject to it have a legitimate complaint to lodge; their situation is not merely bad, or regrettable; it is *wrongful*. Furthermore, to be a member of an unjust society is to bear a connection to that injustice. And yet this connection is not straightforward. By merely belonging to a society in which people are subject to injustice, I do not thereby do that wrong to them; in fact, for certain injustices, there may be no one who can be described as their perpetrator. So, to ask what is required of an individual in the face of social injustice, or in order to avoid to the occurrence of injustice, is to ask a version of the broad question introduced above. That our account of what justice requires of individuals should be premised on a prior understanding of the social institutions in place is not controversial: at least, not amongst writers of a certain theoretical bent. However, as I argue across Chapters 1, 2, and 3, keeping this point in view can help us to avoid certain confusions and tensions.

Whilst my titular question is most explicitly pursued by Part 1, it is also relevant to the discussion of Part 2. Chapter 3 raises the question of whether we might be morally required not to do things which form part of wider behavioural patterns that collectively reproduce social injustice. We can pursue this question, I argue, by asking whether we have a reason to refrain from collectively harmful actions. This is because, at least on one popular way of understanding the notion of harm, social injustice can be thought of as something which is harmful to those who are subject to it. As such, if we have a reason to refrain from collectively harmful actions, then we will have good grounds for thinking that we are required not to act in ways which collectively sustain injustice.

I will round off my introductory comments with a synopsis of the opening three chapters. Chapter 1 approaches the question of what justice requires of

individuals by first covering some familiar terrain. I start by introducing the principles which form the basis of the theory of social justice first published 50 years ago by John Rawls. Rawls claims that his principles of justice apply to the basic structure of society: the major institutions which collectively determine how fundamental rights, duties, and benefits get distributed within society. He also claims that this differentiates the principles of justice from other sorts of moral principle which govern the actions of individuals in particular circumstances. This dual claim, which I dub the 'basic structure restriction', has been much discussed in recent years, primarily as a result of two well-known critiques due to Liam Murphy and G. A. Cohen. My ostensive aim in Chapter 1 is to defend Rawls' basic structure restriction against these two criticisms. However, for those who dislike flogging dead horses, my underlying reason for revisiting these well-trodden debates is to develop a picture of what Rawls' principles of justice require of individuals. *Contra* Murphy, I propose that these principles do not merely provide a blueprint for the design of institutions, but in fact make requirements of the individual members of society. However, *contra* Cohen, these requirements need not pervade every area personal life. Private individuals, I argue, are often unable to shape the rules (both legal and conventional) which comprise the major institutions, and so can often be exempted from the requirements implied by the principles of justice. Exactly what is required of an individual by these principles will depend upon her place within existing social institutions.

Chapter 2 pursues my titular question in a more direct way. Here, I focus on the conception of equal opportunity presented within Rawls' theory. This conception, I argue, faces a number of difficulties. However, these can be avoided if we adopt a recent theory of equal opportunity proposed by T. M. Scanlon, a theory which develops Rawls' in several ways. Once Scanlon's theory is in place, we can ask what requirements it makes of individuals (given the basic structure restriction). Understood in this way, I argue that a principle of equal opportunity requires less of individuals than is sometimes thought. More specifically, I argue that it does not conflict with a plausible principle of parental partiality, at least not in certain areas which have often

been thought to be sites of such a conflict. Again, the obligations which Scanlon's conception of equal opportunity implies, and the ways these interact with reasons of parental partiality, depend upon the organisation of various social institutions.

Chapter 3 continues the preceding two chapters' focus on social injustice and our responsibilities with respect to it. Here, however, my primary subject is not Rawls' theory of justice, but rather the notion of structural injustice: or, more specifically, the account of this notion provided by Iris Marion Young. Young's work on structural injustice has had a very great influence on contemporary political theory. Yet I confess that I struggle to retain a clear grasp of the concept. Chapter 3 tries to rectify this. My aims are two: first, to reach a satisfactory understanding of the notion of structural injustice; and second, to establish what is gained by introducing that concept, so understood, into a theory of social justice. The resulting view draws us back to the idea that the duties to which we are subject can depend upon the social institutions within which we find ourselves. At the centre of Young's project is a theory which assigns responsibilities to individuals on the basis of the causal roles they play in sustaining structural injustice. I reject this theory. It is important for theorists of justice to recognise that injustice can be structural, I argue, not because this reveals the extent of our causal entanglement with injustice, thereby expanding what is required of us in its presence. Rather, it is important because it reveals the extent to which social institutions pervade our lives and influence our prospects. This is all the more reason to think that our first responsibility with respect social justice is to shape these institutions for the better; and precisely what this will require of a given person will depend upon how she is positioned within them, and how they are arranged.

In 2014, during my first term as an MPhil student at UCL, I enrolled on two courses: one, led by Jo Wolff, on the philosophy of Iris Marion Young; and another, led James Wilson, on the ethics of climate change. Together, these courses laid the foundations for the project which I subsequently embarked upon, a project on which this thesis is a kind of status report. I was captivated by Young's work. I had not before encountered serious political philosophy

which took as its starting point a grassroots concern with social justice; and, for this reason, I considered Young's thinking to be deeply important. At the same time, I remember being irked by a paper of Walter Sinnott-Armstrong's. In that paper, Sinnott-Armstrong argues that frivolous actions which make small, causal contributions to the climate crisis, such as a Sunday joyride in a gas-guzzling SUV, are morally permissible: since, on its own, such an action could not emit enough greenhouse gases to make any difference to global temperatures. I believed these arguments to be seriously wrongheaded. Moreover, as I explained in a term paper, I believed that Young's work on structural injustice could be applied to the climate crisis to show that, whilst they were not necessarily to blame for their behaviour, individuals whose actions made causal contributions to the harms of climate change had a responsibility to join with others to change existing social structures.

As the foregoing paragraphs attest, I have since changed my mind. I now hold a view on collective harm which is roughly continuous with Sinnott-Armstrong's (though I am less inclined to affirm his empirical claims about the harmlessness of greenhouse-gas-emitting actions). And I am no longer so enamoured with Young's work on structural injustice. I leave it to my reader to judge the sagacity of this change in view.

Part I

The Demands of Justice

1

The Basic Structure Restriction

This chapter concentrates on an element of John Rawls' theory of social justice which I refer to as the 'basic structure restriction'. This consists of two claims: first, that the principles of justice apply primarily to the basic structure of society; and second, that this differentiates them from other sorts of moral principle. My aims are two. My immediate aim is to defend Rawls' dual thesis against two well-known challenges: one pressed by Liam Murphy, and another due to G. A. Cohen. However, I also have a less explicit, underlying purpose: to build up an account of what Rawls' principles of justice require of individual members of society. On the picture which emerges, these principles require us to act so that the societal distribution of fundamental rights, duties, and benefits determined by the major social institutions accords with that prescribed by the principles. The argument proceeds as follows: Section 1 elaborates Rawls' basic structure restriction in further detail; Section 2 addresses Murphy's critique; and Section 3 addresses Cohen's.

1. The principles of justice and the basic structure

According to Rawls' theory, a society is more just to the extent that it satisfies his principles of justice. In their definitive statement, these principles are two. The first principle, or, as it is more commonly called, the *liberty principle*, stipulates that:

> Each person is to have an equal right to the most extensive total system of equal basic liberties compatible with a similar system for all (Rawls 1999, p. 266)

Rawls refrains from offering a definitive list of the liberties required under the liberty principle. Principal among them, however, are:

> … political liberty (the right to vote and to hold public office) and freedom of speech and assembly; liberty of conscience and freedom of thought; freedom of the person, which includes freedom from psychological oppression and physical assault and dismemberment (integrity of the person); the right to hold personal property and freedom from arbitrary arrest and seizure as defined by the concept of the rule of law (Rawls 1999, p. 53)

The second principle stipulates that:

> Social and economic inequalities are to be arranged so that they are both:
>
> (a) to the greatest benefit of the least advantaged, consistent with the just savings principle[3], and
>
> (b) attached to offices and positions open to all under conditions of fair equality of opportunity (Rawls 1999, p. 266)

The two clauses of this principle are in fact principles in their own right: clause (a) being known as the *difference principle*, and clause (b) as the *principle of fair equality of opportunity*. As such, I will speak of Rawls' three principles of justice: the liberty principle, the principle of fair equality of opportunity, and the difference principle.

My topic in this chapter will be this: What, if anything, do Rawls' principles of justice demand of individuals? This question arises because of the way in which Rawls characterises the virtue of justice which these principles are intended to define. This virtue, he claims, is not personal, but social. That is to say, the standard defined by the principles of justice is not primarily one to which we should hold individuals, their conduct, decisions, or dispositions, but rather one to which we should hold social systems:

[3] The just savings principle defines the proportion of wealth which must be saved or invested for future generations, in order that "each generation receives its due from its predecessors and does its fair share for those to come" (Rawls 1999, p. 254).

> Many different kinds of things are said to be just and unjust: not only laws, institutions, and social systems, but also particular actions of many kinds, including decisions, judgements, and imputations. We also call the attitudes and dispositions of persons, and persons themselves, just and unjust. Our topic, however, is that of social justice. For us, the primary subject of justice is the basic structure of society, or more exactly, the way in which the major social institutions distribute fundamental rights and duties and determine the division of advantages from social cooperation (Rawls 1999, p. 6)

Later on, at the outset of Chapter 2 of *A Theory of Justice*, Rawls asserts that this earlier claim, that the primary subject of social justice is the basic structure of society, differentiates the principles of justice from other sorts of moral principle:

> The principles of justice for institutions must not be confused with the principles which apply to individuals and their actions in particular circumstances. These two kinds of principles apply to different subjects and must be discussed separately (Rawls 1999, p. 47)

Taken together, the thoughts expressed in these two passages comprise what I will call the *basic structure restriction*. This is composed of two claims: first, that the principles of justice apply primarily to the basic structure of society; and second, that this differentiates them from "the principles which apply to individuals and their actions in particular circumstances".

My principal aim in this chapter is to argue that Rawls' basic structure restriction survives the objections of Murphy and Cohen. These two critiques have had a strong influence both on the subsequent reception of the basic structure restriction and on its interpretation, and they have prompted a wealth

of responses.[4] My reasons for adding to an already extensive literature are two. First, I believe that Murphy misinterprets Rawls' view, a point which has been passed over by other commentors. On a more accurate reading of Rawls' text, I will argue, the basic structure restriction avoids Murphy's principal objection. Second, whilst Cohen's critique has attracted much critical attention, there is a move in the argument which remains relatively unscrutinised: namely, Cohen's claim that if the basic structure of society is taken to include informal social conventions, then the principles of justice must make requirements of private individuals across a range of day-to-day circumstances. I will argue that, when subject to closer critical scrutiny, this claim turns out to rely on an implausible assumption about our ability to shape social conventions. A final disclaimer: since my primary aim is to defend Rawls' basic structure restriction against criticisms levelled against it, I will have little to say about the positive reasons which Rawls gives for endorsing it. I will, however, briefly address this question where it is relevant, and revisit it in Chapter 3.[5]

2. Murphy

Murphy proposes an influential interpretation of Rawls' basic structure restriction, according to which this dual claim implies that the principles of justice themselves do not make any requirements of individuals. Insofar as people have responsibilities with respect to social justice, on Murphy's interpretation, these are given by a separate principle which is defined in terms of the principles of justice. This view, Murphy argues, is implausible for a number of reasons, though primarily for one. In this section, I argue that Murphy misrepresents Rawls' view; Rawls' principles of justice do in fact make requirements of individuals. Moreover, on my preferred reading, the basic structure restriction avoids Murphy's principal criticism.

[4] For responses to both Murphy and Cohen, see (Julius 2003; Pogge 2000; Scheffler 2006). For responses exclusively to Cohen, see (Cohen 2001; Estlund 1998; Neufeld 2009; Schouten 2013; Tan 2004; Williams 1998).

[5] For some helpful discussions of Rawls' reasons for proposing the basic structure restriction, see (Cohen 2001; Scheffler 2006). A. J. Julius offers independent grounds for thinking that "something like [Rawls'] basic structure enjoys an intrinsic claim to special moral attention" (Julius 2003, p. 322).

2.1. Dualism and monism

Murphy ascribes to Rawls a view which he refers to as "dualism", according to which "the two practical problems of institutional design and personal conduct require, at the fundamental level, two different kinds of practical principle" (Murphy 1998, p. 254). He contrasts this view with "monism", which denies "that there could be a plausible fundamental normative principle for the evaluation of legal and other institutions that does not apply in the realm of personal conduct" (Murphy 1998, p. 254).

Defined as such, the dualist position comprises of two claims: first, the claim that the principles of justice are fundamental, that is, that they are not derivable from some further, more basic normative principle; and second, the basic structure restriction, interpreted in a particular way. This complicates the monist position. Murphy tends to present this as a single view. However, as the rejection of dualism, monism in fact covers a disjunction of views; a monist could deny either or both of the claims endorsed by the dualist. That is, a monist could hold any of the following: (a) that the principles of justice are derivative; (b) that they do not apply primarily to the basic structure, and so are indistinct from the principles which apply to individuals; or (c) *both* (a) and (b).

This complication noted, let's move on to consider the principal focus of this subsection: Murphy's treatment of the basic structure restriction. Murphy interprets this idea as a "bifurcation of the normative realm into one set of principles for institutions and another for people" (Murphy 1998, p. 279). The principles of justice, according to the dualist, "are used to evaluate institutions and do not apply to people's own choices" (Murphy 1998, p. 271). I read these claims as follows. We can distinguish between two sorts of normative principle: *personal* and *impersonal*. Personal principles specify that, under certain circumstances, a person ought to act in a particular way. These are the principles for individuals, those "which apply to individuals and their actions in particular circumstances" (Rawls 1999, p. 47). By contrast, impersonal principles merely state that things ought to be a certain way; rather than stipulating how people should act, they stipulate how the world should be.

According to the dualist, the principles for institutions are impersonal; they specify how social institutions ought to be, without specifying what anyone ought to do.[6]

The dualist's "bifurcation of the normative realm" thus provides an interpretation of both components of the basic structure restriction. It provides an interpretation of the claim that the primary subject of the principles of justice is the basic structure. According to Murphy, to endorse this claim is to affirm that the principles of justice are impersonal in content: that they stipulate how the major institutions of society ought to be, without stipulating how anyone ought to act. Furthermore, it is evident how this distinguishes the principles of justice from the principles for individuals. Whilst the content of the former is impersonal, the latter are personal in content: they do not merely stipulate how things ought to be, but specify how individuals ought to behave.

As Murphy (1998, p. 270) points out, however, there is an immediate difficulty with claiming that the content of the principles of justice is impersonal. These principles will imply that social institutions ought to be organised in a certain way. But, since they are impersonal, they will not specify that anyone ought to organise them in that way. So, if the normative principles which concern social institutions are exhausted by the principles of justice, then there will be no normative pressure in the direction of realising their prescriptions; if social institutions fall short of the standard defined by the principles of justice, then things will not be as they should, but no one will have failed to act as they should. Such a view might well be consistent. But, as a theory of social justice, it seems unattractive. For Rawls at least, the principles of justice are to guide both criticism and *reform* of social institutions (see e.g. Rawls 1999, pp. 11-12).

[6] I am not suggesting that the distinction between personal and impersonal principles is either exclusive or exhaustive. On exclusiveness, consider the principle: things ought to be such that I keep my promises. Is this personal, impersonal, or both? I leave it open. On exhaustiveness, consider the principle: promises ought to be kept. This principle doesn't seem to meet the criteria for being personal, since it doesn't specify an agent. But neither does it seem to meet the criteria for being impersonal, since it states that actions of a certain sort ought to be performed. The point of the distinction is simply to demonstrate that whilst some normative principles stipulate that someone ought to do something, others need not.

What the dualist needs, then, is a personal principle which instructs individuals to establish, maintain, or reform social institutions so that they satisfy the impersonal prescriptions of the principles of justice. In Rawls' theory, Murphy (1998, p. 271) claims, this lacuna is filled by the natural duty of justice: a personal principle which directs individuals to support and comply with just institutions when they exist, and to further their establishment when they do not (Rawls 1999, p. 99).

Once the dualist view has been supplemented by the duty of justice, Murphy claims that we can rephrase the disagreement between dualism and monism in terms of the responsibilities which they assign to individuals:

> Whereas monism holds that people have direct responsibility for justice, dualism holds that as far as justice is concerned, the responsibility of people is mediated by institutions. If just institutions must aim at equality, monism holds that people must aim at equality too; dualism holds, by contrast, that people must aim at the existence of institutions that aim at equality (Murphy 1998, p. 271)

According to dualism, individuals' responsibilities with respect to social justice are defined by the duty of justice; and this principle directs us to ensure that the impersonal prescriptions of the principles of justice are instigated. We are to ensure, that is, that society's major institutions are as the principles of justice stipulate they ought to be. Monism, by contrast, does not need to be supplemented by the duty of justice. According to this view, the principles of justice are already personal in content. As such, there is no need for an additional personal principle to define the responsibilities of individuals with respect to social justice: these responsibilities are already laid out by the principles of justice themselves.

Notice that, given the complication I introduced a few paragraphs ago, this contrast between the monist and dualist positions is not entirely accurate. A

monist could in fact endorse the basic structure restriction, and instead reject the other dualist claim: the claim that the principles of justice are fundamental. Assuming Murphy's interpretation of the basic structure restriction, this would yield the following view: the principles of justice are impersonal in content, and are derivable from some further, more basic normative principle. Would a monist view of this sort imply a "direct responsibility for justice", or would it need supplementing by Rawls' natural duty of justice? It might not require supplementation if the more basic normative principle from which the impersonal principles of justice are derived was itself personal. In this case, the more basic personal principle might direct individuals to act in ways which satisfy the impersonal principles which it implies. However, if the principles of justice are derivable from a fundamental impersonal principle, then they would require supplementation by Rawls' duty of justice. Thus, if normative principles can be both impersonal and fundamental, then there is logical space for a monist position which does not imply a "direct responsibility for justice".

However, for the sake of argument, let's sidestep this complication. I will assume that the kind of monist we are concerned with is the kind which denies the basic structure restriction, as interpreted by Murphy: that is, the kind which holds that the principles of justice are personal in content. Given this assumption, Murphy's contrast stands: the dualist holds that an individual's responsibilities with respect to social justice are defined by the duty of justice, whilst the monist holds that these responsibilities are already given by the principles of justice themselves.

This contrast forms the basis of Murphy's principal objection to dualism.[7] This objection asks us to compare what dualism and monism would require of individuals under conditions of injustice. Monism, Murphy claims, would require individuals to combat and remedy injustice directly: since monism (or, at least, the kind of monism we are concerned with here) implies that the

[7] Murphy presents a number of arguments which either attack dualism directly, or attack arguments designed to support it. However, he believes the one on which I focus in the text to be the strongest.

principles of justice themselves direct individuals to act. For instance, suppose that there are social and economic inequalities within a given society which do not satisfy the difference principle, and that this can be remedied without infractions of either the principle of fair equality of opportunity or the liberty principle. On the monist view, the difference principle itself would require members of this society to eliminate these inequalities, perhaps by requiring some of the most advantaged individuals to transfer portions of their wealth to the least advantaged. By contrast, according to dualism, the difference principle would not require individuals to eliminate unjust inequalities. Rather, their responsibilities would be given by the duty of justice, which directs them to support or establish social institutions the purpose of which is to combat injustice.

Murphy argues that the requirements which dualism makes of individuals under conditions of injustice are sometimes implausible. This is the case when an individual can mitigate injustice more effectively by doing so directly, rather than by supporting institutions designed to promote justice. For instance, suppose that an individual is one of the most advantaged individuals in a distribution of wealth which does not satisfy the difference principle. It might turn out that this individual would be able to mitigate injustice more effectively by simply transferring wealth to the least well off, rather than by, for example, donating their money to a political campaign for egalitarian redistribution. In this case, monism implies that the individual ought to transfer a portion of their wealth directly to the worst off; this is required of them by the difference principle. Dualism, on the other hand, implies that they ought to choose a less effective means of mitigating injustice: since their responsibilities in this regard are given by the duty of justice, and this requires them to support just institutions. In such cases, Murphy claims, the dualist requirement to choose the latter option is irrational. Requiring individuals to support just institutions under such conditions amounts to requiring individuals to promote an end by choosing worse means:

> [according to dualism] even if the individual could do more to reduce inequality, alleviate suffering, or whatever, by direct

> action, this is not what justice requires her to do. Justice requires her to promote just institutions even if she is sure that the aim of the just institutions she is promoting would be better served if she herself pursued that aim directly. How could this be right? (Murphy 1998, p. 281)

2.2. Reinterpreting the basic structure restriction

In the preceding subsection, I presented Murphy's interpretation of the basic structure restriction. On Murphy's reading, when Rawls says that the principles of justice apply primarily to the basic structure, we should take this to mean that the content of these principles is impersonal: they stipulate how the major social institutions of society ought to be, without specifying that anyone ought to make them so. This feature differentiates the principles of justice from the principles for individuals, since the latter are personal in content. In this subsection, I will argue that this is not a plausible interpretation of Rawls' position; Rawls does not think that his principles of justice are impersonal. In its place, I offer my own reading of the basic structure restriction. This reading, I argue, is superior to Murphy's on two counts: first, it coheres better with Rawls' text; and second, it avoids Murphy's objection.

That Rawls does not intend for the principles of justice to be impersonal in content can be inferred from various points in *A Theory of Justice*. For instance, there is the long discussion, comprising Chapter 8, of how a society ordered around the principles of justice would foster a "sense of justice" amongst its members: "a normally effective desire to apply and act upon the principles of justice, at least to a certain minimum degree" (Rawls 1999, p. 442). If a person can desire to act on the principles of justice, then those principles must require things of us.

However, I suggest that Murphy's error is most clearly revealed by considering the argument which Rawls offers for thinking that the parties in the original position would choose to be bound by the natural duty of justice (Rawls 1999, pp. 293-295). Rawls sets the ground for this argument by

making two simplifying assumptions. First, he assumes that his principles of justice have already been chosen to govern the basic structure of society. Second, he assumes that the choice facing the parties is between the natural duty of justice, and the principle of utility. As such, the decision problem to be resolved is this: given that Rawls' principles of justice are to govern the basic structure, would it be more rational for the parties to choose the duty of justice to define their institutional responsibilities, or the principle of utility?

Rawls' argument then proceeds as follows. If the parties were to adopt the principle of utility, he claims, then this choice would lead to a conflict of requirements. The parties, it is assumed, have already chosen the principles of justice to regulate the basic structure. But: "[t]he principles for institutions have … consequences for the acts of persons holding positions in these arrangements" (Rawls 1999, pp. 294-295). Rawls offers two examples of a situation in which the principles of justice imply requirements on individuals: that of a citizen deciding how to vote between political parties, and that of a legislator deciding whether or not to favour a certain statute (Rawls 1999, p. 294). In such cases, he claims, a person ought to support the party, or favour the statute, which best furthers the principles of justice. However, the principle of utility would require an individual to support that party or statute which maximises (aggregate or average) utility. Since the acts of government which best satisfy Rawls' three principles of justice will not always maximise either aggregate or average utility, choosing the principle of utility to define people's institutional responsibilities would thereby lead to inconsistency. "To avoid this conflict" Rawls writes "it is necessary, at least when the individual holds an institutional position, to choose a principle that matches in some suitable way the two principles of justice" (Rawls 1999, p. 295).

As we saw in the preceding subsection, Murphy interprets Rawls' introduction of the natural duty of justice as grounded in the need to supplement the impersonal principles of justice with a personal principle; since the principles of justice do not make requirements of individuals, they must be accompanied by some further principle which does. However, we can now see that this interpretation gets Rawls' reasoning back-to-front. In

fact, the argument for introducing the duty of justice *relies* on the assumption that the principles of justice imply requirements on individuals. Far from being purely impersonal in content, these principles direct individuals to take steps to ensure that their prescriptions are met: to vote for political parties whose manifestos best conform to the principles of justice, to urge others to vote likewise, to use one's office to support reform in the direction of the principles' requirements, and so on (Rawls 1999, p. 294).

Murphy's interpretation of the basic structure restriction will not do, then, as a reading of the text. When Rawls says that the primary subject of the principles of justice is the basic structure, he is not claiming that they stipulate merely how the major social institutions ought to be, without stipulating how people ought to act. As such, this cannot be what differentiates the principles of justice from the principles for individuals. How, then, should we read these claims?

I propose that we interpret the basic structure restriction as follows. In claiming that the principles of justice apply primarily to the basic structure, Rawls is not denying that they make requirements of individuals, but is rather claiming that they make requirements only of a restricted range of actions. These are those actions which can influence the basic structure of society: that is, actions which are capable of affecting the distribution of fundamental rights, duties, and advantages determined collectively by society's major institutions. This explains Rawls' focus on institutional positions when giving examples of the sorts of situation in which the principles of justice imply requirements on individuals. A person endowed with legislative powers, for instance, is in a position to influence the basic structure of society; by proposing and supporting certain statutes, they can shape the distribution of rights, duties, and benefits determined collectively by the major institutions. Or again, when an individual casts their ballot in a general election or a referendum, there is a chance that their vote will affect the basic structure.[8]

[8] For an extended discussion of voting, see Chapter 5 (Section 5.3).

This provides an interpretation of the first element of the basic structure restriction: the claim that the primary subject of the principles of justice is the basic structure. But we still need an interpretation of the second: the claim that the principles of justice can be differentiated from the principles "which apply to individuals and their actions in particular circumstances" (Rawls 1999, p. 47) on the basis that they apply to different subjects. My interpretation of the first component might seem to mystify the second: since, on my reading, the principles of justice require certain actions of individuals in particular circumstances. However, I suggest that it can be made sense of. The difference between the principles of justice and the principles for individuals, I propose, is one of scope. The principles for individuals will be relevant across a wide range of circumstances. They will include, for instance, the duties of non-maleficence, beneficence, fidelity, etc. Such principles make requirements of us in a broad variety of possible situations; it is quite common, for instance, that we will be in a position to harm someone else. By contrast, it is rarer that an individual will be in a position to affect the basic structure of society.[9] As such, the circumstances in which the principles of justice make requirements on us will be less numerous than those in which the principles for individuals apply.

This, then, is my favoured interpretation of the basic structure restriction: the principles of justice require us to act in ways which will cause the basic structure of society to align closer with the specifications of those principles; and this, moreover, differentiates these principles from the principles for individuals, since the latter will apply across a wider range of circumstances. This interpretation might alternatively be formulated in terms of the reasons which these two types of principle, the principles of justice and those for individuals, pick out. Moral principles isolate reasons for action: they identify a particular reason for acting in some way or other, and stipulate that it can often be sufficient to defeat other countervailing reasons.[10] The principles of

[9] I will consider a challenge to this claim in Section 3.2.

[10] Rawls sketches such a conception of moral principles when he writes:

> A principle taken alone does not express a universal statement which always suffices to establish how we should act when the conditions of the antecedent

justice are no different. To say that these principles apply to the basic structure is not to deny this, but rather to claim that the reasons which they isolate are exemplified only in certain circumstances: namely, those in which our actions can influence the basic structure.

This way of reading the basic structure restriction coheres with the passages of *A Theory of Justice* which I highlighted above, in which Rawls seems to emphasise that the principles of justice can make requirements of individuals. As such, it is a more accurate reading of Rawls' text than Murphy's interpretation. But its merits are not merely scholarly. My interpretation of the basic structure restriction also avoids Murphy's principal objection to dualism. This objection arises, recall, in situations in which acting on Rawls' natural duty of justice would be a less effective way for an individual to mitigate injustice. In such cases, Murphy claims, dualism would require individuals to pursue this less effective means, since it holds that the duty of justice exhausts our responsibilities with respect to social justice. However, given my interpretation, Rawls does not need to say this. If the basic structure of society fails, in some way, to meet the specifications of the principles of justice, and an individual is in a position to change that for the better, then, on my reading, the principles of justice require them to do so. And if it turns out that they can mitigate injustice more effectively by, say, transferring a portion of their wealth to the worst off rather than by promoting some institutional policy, then they are required to choose the more effective means.

At this point, one might raise a worry: 'If the principles of justice themselves make requirements of individuals, then doesn't the duty of justice become redundant? Why would the parties in the original position choose to be bound by this duty, if the requirements it makes of them are already implied by the

---

> are fulfilled. Rather, first principles single out relevant features of moral situations such that the exemplification of these features lends support to, provides a reason for making, a certain ethical judgement (Rawls 1999, p. 300)

For further development of an account along these lines, see (Scanlon 1998, pp. 197-202).

principles of justice?' But consider more closely what the duty of justice requires:

> This duty has two parts: first, we are to comply with and to do our share in just institutions when they exist and apply to us; and second, we are 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 1999, pp. 293-294)

In certain situations, what this duty requires of an individual will coincide with the directives implied by the principles of justice (as in the institutional cases discussed above). But in others, the duty of justice will make requirements which are not implied by the principles for institutions. For instance, take the requirement to comply with the rules of just institutions. It could be that by breaking some institutional rule, an individual would not cause the distribution of rights, duties, and benefits determined collectively by the major institutions to move further away from that required by the principles of justice; in breaking this rule, the individual would not prevent the institution from performing its function, nor would they cause general non-conformity which would threaten the institution's stability. As such, the principles of justice themselves will not direct this individual to comply with the rule in question: at least, not on this particular occasion. However, as formulated above, the natural duty of justice does require this.

2.3. The difference principle

The preceding subsection offered a new interpretation of the basic structure restriction, according to which the principles of justice require us to shape the basic structure of society so that it better conforms with their specifications. In this subsection, I address a difficulty which this interpretation faces in relation to one of the principles of justice in particular: the difference principle.

Rawls writes that the difference principle:

> … applies to the announced system of public law and statutes and not to particular transactions or distributions, nor to the decisions of individuals and associations, but rather to the institutional background against which these transactions and decisions take place (Rawls 1993, p. 283)

Here, Rawls seems to be saying that the difference principle should not be used as a guide for individual economic choices; when making decisions about how to spend their money, individuals and associations should not aim to ensure that any inequalities which their transactions generate are most beneficial to the worst off. One might think that this causes problems for my interpretation of the basic structure restriction. My view claims that the principles of justice make requirements of individuals. But if the difference principle cannot serve as a guide for individual choices, one might press, then how can it make requirements of us?

Moreover, these particular observations about the difference principle might suggest a rival interpretation of the basic structure restriction. In claiming that the difference principle should not be used as a guide for individual choice, one might propose, Rawls is simply reiterating his claim that, like the other two principles of justice, this principle applies primarily to the basic structure. On this view, to say that the subject of the principles of justice is basic structure is, at least in part, to say that they are unsuitable to be used as guides for individual choice. This differentiates these principles from the principles for individuals, as the latter can and should be used as decision-making guides. This interpretation of the basic structure restriction will be inconsistent with my own if, as suggested above, a principle which cannot serve as a guide for individual choice cannot make requirements of individuals. Indeed, if this is correct, then these considerations will evidence Murphy's view that, according to Rawls, the principles of justice are impersonal in content.

Before addressing these challenges directly, it will be instructive to consider the context in which the quoted passage occurs. What moves Rawls to say

that the difference principle does not apply to individual transactions? Why does he deny that individuals should use this principle to guide their day-to-day economic decisions? Rawls' answer is that such a requirement would be unfeasible. It would require individuals to foresee the ramifications which the totality of transactions will have collectively on the societal distribution of income and wealth, and continually to adjust their own transactions accordingly (Rawls 1993, p. 268). Such foresight is clearly beyond the powers of a typical human being.[11]

The unfeasibility of requiring individuals to monitor their own economic activities so as to ensure a just distribution of income and wealth leads Rawls to posit the need for what he calls an "institutional division of labour" (Rawls 1993, p. 268). This consists, essentially, in a division between two sorts of legal regulation. The first sort would direct individuals to avoid certain wrongs in their economic dealings with one another: wrongs such as fraud, duress, exploitation etc. The law of contracts, and certain other areas of private law, would fall under this first category of regulation. The second sort of rule would be designed to ensure that these various individual interactions collectively result in a just distribution of income and wealth. But, unlike the first sort of regulation, it would not achieve this by instructing individuals to pursue that end directly. Rather, it would require them to do certain other things which indirectly produce the desired distribution. Income, inheritance, and property tax, for instance, would fall under this second category. Such forms of regulation direct individuals to relinquish some portion of their holdings. But, under the institutional division of labour, the end which this requirement is designed to achieve is not the relinquishment itself, but rather the distribution of goods which results once the totality of transactions are taken into account.[12]

[11] Here, I follow Samuel Scheffler's (2006) interpretation of Rawls' position.

[12] Seana Shiffrin (2000, pp. 234-236) has argued that the unconscionability doctrine, which enables courts to refuse to uphold contracts which they deem to be exploitative, can be used as a tool to achieve egalitarian redistributive aims. If correct, Shiffrin's point demonstrates that the aims of contract law and those of the taxation system cannot be quite as neatly distinguished as I suggested in the text. However, I submit that it is consistent with the spirit of Rawls' institutional division of labour. First, it is not clear that the unconscionability doctrine directs members of the public to act in any particular way. Rather, it permits courts to act in a certain way: namely, to refuse to uphold exploitative contracts. As such, the

Let's return now to the challenges facing my interpretation of the basic structure restriction. These challenges are two, but they both pivot on a common assumption: that if a principle is unsuitable to serve as a guide for individual decision making, then it cannot make requirements of individuals. The first challenge is local to the difference principle. It says that, since the difference principle cannot be used as a decision-making guide, it cannot, *contra* my interpretation, require individuals to act. The second challenge proposes a rival interpretation of the basic structure restriction which supports Murphy's. On this interpretation, the claim that the principles of justice apply to the basic structure implies that they cannot serve as guides for individual decision making; and that this, in turn, implies that they do not make requirements of individuals.

Rawls' comments on the institutional division of labour, I suggest, cause difficulties for the second challenge. More specifically, they suggest that not all of his three principles of justice would be unsuitable to serve as decision-making guides. Rawls' reason for denying that the difference principle should serve as such is that it would be impracticable for it to do so. Yet it is not clear that it would be unfeasible for the liberty principle, for instance, to function as a decision-making guide. Used as such, this principle would instruct individuals to ensure that their actions do not result in a societal distribution of rights and duties which denies some people their equal basic liberties. Complying with this instruction does not seem to require the same kind of inhuman foresight involved in using the difference principle as a guide for individual choice. As such, it is not obvious that Rawls would deny that the liberty principle can be used as a decision-making guide; or at least, it is not clear that he would deny this on the same grounds as he denies that the difference principle can serve as such.

---

doctrine does not obviously fall under the first category of regulation described in the text. Second, even if the unconscionability doctrine could be interpreted as issuing a directive to private individuals, it would not direct them to ensure that their economic dealings result in a just societal distribution of wealth. As such, the doctrine could only be used to achieve redistributive aims in an indirect way.

But if this is correct, then the claim that the principles of justice apply primarily to the basic structure will not imply that they cannot serve as guides for individual choice. Rawls claims that all three of his principles of justice apply to the basic structure. So, if only some of them are unsuitable to serve as decision-making guides, then the implication proposed by the rival interpretation of the basic structure restriction is invalid. That is, it must be possible for a principle both to apply to the basic structure and to serve as a guide for individual decisions.

These are (perhaps not decisive) considerations against the rival interpretation of the basic structure restriction. However, I suggest that there is a more fundamental difficulty facing both of the two challenges to my view. Both of these challenges rely on the assumption that a principle cannot make requirements of individuals if it is unsuitable to serve as a guide for individual decision making. This assumption, I propose, is mistaken.

To see this, consider for a moment the principle of utility. Utilitarians have often observed that this principle would be self-defeating if individuals were consciously to use it as a guide for their day-to-day behaviour. Due to human limitations similar to those which Rawls highlights when introducing the institutional division of labour, if an individual acts with the explicit intention of maximising utility, then, often, they will fail. Working out which of one's options will maximise utility frequently requires a level of foresight beyond that which most humans possess. Given this, if individuals always attempted to make utilitarian calculations when faced with a choice, we could expect frequent miscalculations. In fact, individuals might make choices which maximise utility more frequently if they deliberated using certain customary rules of thumb, rather than the principle of utility itself.

However, this does not imply that the principle of utility makes no requirements of individuals. Rather, it clarifies what it is that this principle demands of us: the principle requires that individuals comply with it indirectly. If customary conventions lead individuals to make choices which maximise utility, then when individuals use those conventions to guide their

deliberations they will be conforming to the principle of utility; the fact that this is not their primary aim does not foreclose it. Moreover, if by using these customary conventions, individuals will make utility-maximising choices more reliably than they would using the principle of utility itself, then the principle of utility requires that they deliberate using those conventions. Thus, in saying that their principle should not be used as a maxim of individual choice, utilitarians are not denying that it makes demands on individuals. Quite the opposite; they are explaining what the principle of utility requires of us.

What this shows is that the assumption made by both challenges to my position is false; a principle can make requirements of individuals, whilst being unsuitable to serve as a guide for individual choices. Indeed, the principle of utility might itself require us not to use it as a decision-making guide. Combining this thought with my interpretation of the basic structure restriction would give us the following view. The difference principle requires us to shape the basic structure so that the resulting distribution of wealth is to the greatest benefit of the least well off. However, if we were to organise our day-to-day transactions with the aim of satisfying this requirement, then we would most likely fail. As such, the difference principle requires us to comply with it indirectly. That is, we shouldn't attempt to comply with the difference principle by monitoring our daily transactions to ensure that, given the totality of transactions which others will make, the resulting distribution of wealth is most advantageous to the least well off. Rather, we should aim to comply with a taxation system which is itself designed to ensure a just distribution of income and wealth.[13]

[13] This is consistent with Rawls' publicity condition on the principles of justice, according to which it can be assumed that "everyone will know about these principles all that he would know if their acceptance were the result of an agreement" (Rawls 1999, p. 115). What I am suggesting is that the difference principle requires individuals to comply indirectly with its directive to shape the basic structure so that socio-economic inequalities are most beneficial to the worst off. Individuals do not need to be ignorant of that directive in order to avoid complying with it directly. It is sufficient that they believe that attempts at direct compliance would likely fail.

Notice that this view does not preclude there being circumstances in which we should comply with the difference principle directly. There may be situations in which an individual can be sure that, by spending their money in one way rather than another, they will bring about a more just distribution of wealth. As such, the present view is consistent with the response I gave in Section 2.2 to Murphy's objection to dualism: if it turns out that an individual can mitigate injustice most effectively by making a one-off transfer of money to the worst off, then the difference principle requires them to do so. The point is simply that, in general, if individuals were to organise their day-to-day spending around the conscious aim of furthering the difference principle, then the resulting distribution of wealth would most likely not be to the greatest advantage of the least well off.

3. Cohen

Murphy's principal objection to the basic structure restriction can, then, be avoided by adopting a more accurate reading of Rawls' text. However, there is another well-known critique of Rawls' idea which cannot be avoided in this way. This critique is due to G. A. Cohen. My aim in this section is to defend the basic structure restriction, interpreted as set out in Section 2.2, against Cohen's criticisms.

3.1. The basic structure objection

Cohen's broader argument is intended as a critique of the invocation of economic incentives to justify inequality. At least on certain interpretations[14], the difference principle will permit inequalities of income if paying certain individuals more than others motivates them to do work which will benefit the worst off. If we suppose that, without financial incentives, these individuals would not work as hard, or would choose to do work which is less beneficial to others, then it will be necessary to pay these incentives in order make the worst off as well off as possible. Cohen points out, however, that this argument assumes that people's preferences and motivations are to be

[14] See (Cohen 2008, pp. 68-69) for the distinction between 'strict' and 'lax' interpretations of the difference principle.

taken as fixed. If, by contrast, the talented[15] were to internalise the difference principle, then their motivations would be such that financial incentives would not be needed in order to maximise the prospects of the worst off; the talented would sincerely believe that economic inequality is admissible only if it is most beneficial to the least well off, and so would prefer to work for the benefit of the worst off without receiving inequality-generating incentives.[16]

Cohen then imagines a possible Rawlsian reply. Individuals who endorsed the difference principle need not, the reply goes, be motivated to work for the benefit of the least well off. This is because, like all of Rawls' principles of justice, the difference principle applies to the basic structure. As such, it does not direct individuals to work in professions which procure the greatest advantages for the least well off; such personal decisions fall outside of the principle's restricted scope (Cohen 2008, pp. 124-125). Cohen dubs this the *basic structure objection*.

Incidentally, my comments in Section 2.3 make possible an alternative reply to Cohen's argument. Cohen imagines that a supporter of Rawls' incentives-based justification of economic inequality would appeal to the basic structure restriction. But, given what I said in Section 2.3, they could alternatively appeal to the claim that the difference principle is unsuited to serve as a guide for individual economic decisions. A reply along these lines need not deny that the difference principle requires individuals to engage in work which benefits the worst off. Rather, it would deny that individuals should

[15] On the sense in which I am using the term, to say that someone is 'talented' is to say that their capacities command a high market value. It implies nothing about the substantive value of those capacities, nor about how that person came to hold them. This is the sense in which Cohen employs the term:

> All that need be true of [the talented] is that they are so positioned that, happily for them, they do command a high salary and they can vary their productivity according to exactly how high it is. But … their happy position could be due to circumstances that are entirely accidental, relative to whatever kind of natural or even socially induced endowment they possess (Cohen 2008, p. 120).

[16] Versions of this argument reappear across a number of Cohen's writings, see (Cohen 1992; Cohen 1997; Cohen 2000, chap. 8; Cohen 2008, chap. 1). My references will be to the most recent of these.

consciously bear this requirement in mind when deciding which jobs to apply for, or when bargaining with their employer over their salary. Instead of using the difference principle to guide their work-related choices, individuals should aim to follow simpler, more localised guidance which is itself designed to ensure that people's behaviour indirectly results in a just distribution of wealth. Such guidance could, in principle, involve financial incentives.[17]

However, let's put this alternative reply to one side. Cohen's response to the basic structure objection is of interest in itself, since it reveals a powerful challenge to the basic structure restriction as interpreted in Section 2.2.

3.2. The fatal ambiguity

Cohen's response to the basic structure objection is organised around an ambiguity which he claims is present in Rawls' characterisation of the basic structure. Recall that Rawls (1999, pp. 6-7) initially describes the basic structure as "the way in which the major social institutions distribute fundamental rights and duties and determine the division of advantages from social cooperation". Cohen (2008, pp. 132-134), however, argues that this description is ambiguous. A society's "major institutions", he claims, might be understood to equate to its legal system, or at least those parts of it which distribute fundamental rights, duties, and benefits. Call this the *juristic reading*. Alternatively, thinks Cohen, the basic structure could be interpreted more inclusively to cover not only a society's legal system, but also "conventions and usages" (Cohen 2008, p. 125), or "(legally optional) social practices" (Cohen 2008, p. 137): that is, schemes of non-legal rules, such as conventions, traditions, usages, customs etc. which prevail in a society. Call this the *conventionalist reading*.[18]

[17] This is not to say that I endorse all uses of economic incentives. For arguments that Rawls' theory does not license unrestricted use of economic incentives, see (Cohen 2001; Scheffler 2006; Tan 2004). For a separate argument against economic incentives inspired by both (G. A.) Cohen and Rawls, see (Shiffrin 2010).

[18] Cohen's own term for the legal aspects of the basic structure is "coercive" structure; and he refers to those non-legal aspects delineated under what I am calling the conventionalist reading as "non-coercive" structure. I have chosen not to follow Cohen's terminology because the presence or absence of coercion does not seem to track what he is primarily interested in: namely, the presence or absence of law. The law need not always be coercive

Cohen's next move is to ask why, on either of these readings, Rawls' claim that the principles of justice apply to the basic structure would imply that they do not apply to the personal sorts of decision discussed above: decisions about which professions to work in, or what salaries to ask for, and so on. According to the interpretation presented in Section 2.2, recall, to say that the principles of justice apply to the basic structure is to say that they require us to shape the basic structure so that it better conforms to their stipulations. Happily (or unhappily?), Cohen's discussion largely coheres with this interpretation. As such, I will assume it in what follows.

Consider first Cohen's juristic reading of the basic structure. Given this reading, would the basic structure objection succeed? That is, would the assumption that the principles of justice apply to the basic structure imply that they do not require us to choose work which benefits the worst off? Cohen concedes that if we adopt the juristic reading, then this inference is valid. This is because he (provisionally) grants that the legal system "arises independently of people's quotidian choices: it is formed by those specialized choices that legislate the law of the land" (Cohen 2008, p. 135).[19] That is, given the juristic reading, the decisions which are capable of affecting the basic structure are limited to a special class: namely, legislative decisions. We might widen this class slightly to include decisions which affect who the legislators are, and what their agenda is. Given this, voting in general elections and referenda, for instance, would count as an action which can influence the basic structure. But even so, the vast majority of day-to-day decisions which private individuals face will not be capable of affecting the basic structure; and decisions about where to work, or what salaries to bargain for, will fall under this second category. Given this, on the juristic reading, such decisions will fall outside of the restricted scope of the principles of justice.

(see e.g. Hart 2012, pp. 26-49, 199-200); and someone could coerce others to comply with non-legal rules.

[19] Cohen (2008, pp. 144-146) later questions this supposition. However, his objection relies on his treatment of the conventionalist reading; and this is precisely what I will call into question in Section 3.3.

So much for the juristic reading. But what about the conventionalist reading? Cohen argues that if we adopt this interpretation, then the basic structure objection fails. Restricting the application of the principles of justice to the basic structure, on its conventionalist reading, will not imply that they fail to apply to quotidian choices. This, he claims, is because social customs are "bound up with the choices that people customarily make" (Cohen 2008, p. 134). As Cohen himself is quick to acknowledge, his meaning here is not clear. He could be taken to be claiming that, on the conventionalist reading, people's customary choices are a *part* of the basic structure. Yet Cohen denies this, asserting that "[a]ctions are … no part of the basic structure, because a structure … is a set of rules" (Cohen 2008, p. 149). How, then, should we understand his line of thought?

Cohen clarifies his claim as follows:

> With respect to coercive [i.e. legal[20]] structure, one may perhaps fairly readily distinguish the choices that institute and sustain a structure from the choices which occur within it. But with respect to informal [i.e. non-legal] structure, that distinction, though conceptually intelligible, is compromised extensionally: when A chooses to conform to the prevailing usages, the pressure on B to do so is reinforced, and no such pressure exists, the very usages themselves do not exist, in the absence of conformity to them … Just as you can ask whether legislators act justly when they create a certain coercive structure, so you can assess for their justice the deliberate daily sustaining acts of the informal structure in which its participants engage (Cohen 2008, p. 135)

His point here seems to be this. In the case of the legal system, we can distinguish between, on the one hand, those choices which determine the constitution of that system of rules, and, on the other, those which consist in

[20] See note 18.

following those rules. As such, given the juristic reading, the basic structure restriction entails that the principles of justice apply only to the former, juristic set of choices, and not to the latter, quotidian set. However, in the case non-legal social rules, this distinction is "compromised extensionally". That is, the choices that determine which informal conventions prevail in a given context are precisely the same as those which consist in following those conventions: these two categories possess the same extensions. When an individual chooses to conform to a given convention, their choice pressurises others to conform as well; and when they choose to transgress, this will influence others to do likewise. But, since there are no formal conditions which stipulate how such rules may be brought into, or out of, existence, informal conventions exist only insofar as people act in accordance with them. So, in choosing to conform to, or transgress, a given convention, one is also making a choice which affects whether or not that convention exists.

If this is right, then, on the conventionalist reading, the range of circumstances in which our actions can influence the basic structure will be much broader than it is on the juristic reading. Any action which either conforms to or transgresses some social convention will be one which can have an effect on the constitution of the basic structure. So, if there are conventions which regulate people's choices about where to work, or what salaries to bargain for, then the principles of justice apply to those choices. They will require individuals to conform to those conventions which are conducive to a just distribution of rights, duties, and benefits, and to transgress those conventions which are not.

We now have a partial reply to the basic structure objection. According to Cohen, if the principles of justice apply primarily to the basic structure of society, then the sorts of work-related choice with which he is concerned will fall outside of their scope only if the juristic reading is correct. An advocate of the basic structure objection must, therefore, endorse the juristic reading.

Cohen's final move is to argue that if we assume that the principles of justice apply to the basic structure, then the juristic reading is implausible. Why

should we think that the principles of justice apply exclusively to a society's legal structure? According to Cohen, Rawls' primary reason for restricting the principles of justice to the basic structure is that it has profound impacts on people's life prospects:[21]

> The basic structure is the primary subject of justice because its effects are so profound from the start. The intuitive notion here is that this structure contains various social positions and that men born into different positions have different expectations of life determined, in part, by the political system as well as by economic and social circumstances (Rawls 1999, p. 7)

Undeniably, the legal system has such effects. But equally undeniable is the fact that non-legal schemes of social convention and tradition can have them as well. As illustration, Cohen (2008, p. 137) asks us to consider conventional gender roles in the family. In many societies, there remains an expectation that, in a heterosexual couple, the woman will take on the bulk of the domestic labour. There is no law which directs women to do this. But the fact that they often do has large and complicated impacts on their relative prospects when compared to those of men.[22]

Since social conventions can have just as profound an impact on people's life prospects as the law, Cohen argues that it would be arbitrary to restrict the application of the principles of justice to the legal system. Given Rawls' own rationale for treating the basic structure as the primary subject of social justice, the application of his principles should be extended to include non-legal social rules, and so he must adopt the conventionalist reading. But if so, then Rawls' principles must apply to a variety of day-to-day choices. Cohen

[21] Scheffler (2006) claims that Cohen overlooks two further reasons which Rawls gives for applying the principles of justice to the basic structure: first, the propensity of society's major institutions to shape people's wants and aspirations; and second, the inadequacy of rules which apply primarily to personal conduct to ensure background justice. Joshua Cohen (2001) and Kok-Chor Tan (2004) also emphasise the first of these two considerations. I have no quarrel with these arguments. They do not, however, affect my own argument of Section 3.3, since there I am exclusively concerned with Cohen's treatment of the conventionalist reading.

[22] See (Schouten 2013, p. 336, fn. 16) for references to empirical studies.

thus presents advocates of the basic structure objection with a dilemma: either it is arbitrary to think that the principles of justice apply primarily to the basic structure, or the scope of those principles extends to include quotidian choices like which jobs to apply for, or how to organise domestic labour (Cohen 2008, pp. 135-137).

Yet Cohen does not stop there. The ambiguity in the notion of the basic structure, he claims, is not only fatal for the basic structure objection. It also wrecks the basic structure's claim to be the primary subject of social justice (Cohen 2008, p. 132). According to Cohen, that claim is either arbitrary or false. It is arbitrary if we suppose that the basic structure is equivalent to the legal system: since the primary reason for restricting the principles of justice to the legal system is also a reason to apply them to social conventions. However, it will be false if we adopt a more expansive view of the basic structure, on which it encompasses social conventions: since taking the basic structure so conceived to be the subject of the principles of justice will imply that they make requirements of individuals across a broad range of quotidian circumstances.

3.3. Controlling conventions

Before considering how Cohen's challenge might be met, let's be clear on the problem facing Rawls' view. Cohen presents his dilemma as a challenge to the claim that the principles of justice apply primarily to the basic structure. However, on reflection, it's not clear that this is the case. On my interpretation, this claim is equivalent to the claim that the principles of justice require us to shape the basic structure so that it better conforms to their specifications. This claim would, it seems, be arbitrary on the first horn of Cohen's dilemma: that is, if we have just as much reason to focus our efforts on prevailing social conventions as we do to focus them on the legal system. However, it's not clear it would be false on Cohen's second horn; the claim that the principles of justice require us to shape the basic structure is not inconsistent with the claim that it makes requirements on us across a wide range of day-to-day situations.

Nevertheless, Cohen's dilemma does pose a problem for the basic structure restriction, as interpreted in Section 2.2. The basic structure restriction, recall, comprises of two claims: first, that the primary subject of the principles of justice is the basic structure; and second, that this differentiates the principles of justice from the principles for individuals. The second horn of Cohen's dilemma, I suggest, creates a problem for my interpretation of the second of these claims. On my view, recall, applying the principles of justice to the basic structure generates a scope distinction between those principles and the principles for individuals; whilst the latter make claims on us across a broad range of situations, the former make claims on us only in a restricted range of circumstances. This distinction, however, is called into question by the second horn of Cohen's dilemma.

So, formulated more precisely, the challenge posed by Cohen's dilemma is this. Either the juristic reading of the basic structure is true, or the conventionalist reading is true. If the juristic reading is true, then the first component of the basic structure restriction is arbitrary. But if the conventionalist reading is true, then the second component is false. Either way, the basic structure restriction, at least as I interpret it, is untenable.

How, then, can this challenge be met? Defenders of Rawls have explored two lines of response to Cohen's dilemma. Kok-Chor Tan (2004, p. 346, fn. 29), Samuel Scheffler (2006) and Blain Neufeld (2009) each independently propose that it might not be arbitrary to endorse the juristic reading and apply the principles of justice exclusively to a society's legal system. If so, then the first horn of the dilemma is acceptable. Alternatively, Andrew Williams (1998) argues that there is a third reading of the basic structure which Cohen overlooks. Interpreted in this third way, limiting the application of the principles of justice to the basic structure is neither arbitrary, nor implies that their scope extends to quotidian decisions. So, if Williams is right, then Cohen's is a false dilemma.[23]

[23] Gina Schouten (2013) has also defended Rawls's restriction of the principles of justice to the basic structure. However, she seems to agree with Cohen that these principles may sometimes require individuals to ensure that their everyday behaviour does not sustain unjust

I will not discuss these responses any further here.[24] My aim in this subsection is to offer an alternative way out of Cohen's dilemma. Unlike the two approaches sketched above, my response focusses on the second horn. More specifically, I focus on Cohen's claim that if the principles of justice direct individuals to shape informal social conventions, then they must apply to quotidian choices. We saw that Cohen's reasoning behind this claim runs as follows: people can influence the constitution of a scheme of informal social rules by choosing to follow or transgress those rules; so, if the principles of justice direct us to shape informal social rules, then they will make requirements of us across a variety of day-to-day situations in which we can choose whether to follow or transgress such rules.

This argument, I suggest, rests on a mistake. In claiming that an individual's conformity/transgression can influence which social conventions prevail in their milieu, Cohen overestimates our capacity to control non-legal social rules. In order to develop this point, it will be helpful to introduce a distinction. One's control over something can be either *direct* or *indirect*. If one possesses direct control over something, then one can bring it about, or induce changes in it, as an immediate effect of one's actions. For instance, I can exercise direct control over whether or not I marry my fiancée simply by uttering or withholding the requisite words at the appropriate moment. By contrast, if someone's control over something is indirect, then they can bring it about, or produce changes in it, only by bringing about certain intermediaries which themselves cause the desired effect. For example, I cannot bring it about that you marry some third party simply by uttering the right words in the appropriate institutional setting. But I might be able to say some things to each of you which will plant certain ideas in your heads, and these ideas may eventually lead you marry one another. I doubt that this distinction is entirely watertight. (I can bring it about that I marry my fiancée

conventions (Schouten 2013, p. 382). As such, her view does not rescue the basic structure restriction, as I interpret it.

[24] For a response to Williams, see Cohen (2008, chap. 8). For criticism of Scheffler and Neufeld, see (Schouten 2013, pp. 366-368).

only by bringing it about that the right sounds come out of my mouth. Does this mean that my control over whether or not I marry turns out to be indirect after all?) But I think it is comprehensible enough for present purposes.

No one, I suggest, has direct control over what is, and what is not, a convention, tradition, usage, more, custom etc. I take this to be an insight of H. L. A. Hart's, and it is nicely illustrated by his recitation of "the story, perhaps apocryphal, that the headmaster of a new English public school announced that, as from the beginning of the next term, it would be a tradition of the school that senior boys should wear a certain dress" (Hart 2012, p. 176). The headmaster's announcement is absurd; and it is absurd precisely because our concept of tradition does not allow that traditions are the sort of rule that could come into, or out of, existence as the immediate result of someone's say-so. Similar things can be said for other sorts of non-legal rule. In this way, such concepts stand in contrast with our concept of law. Certain people do enjoy direct control over the laws which regulate behaviour in their society; indeed, legal systems are continuously revised by the creative and interpretative acts of legislators and judges.

So, if by conforming to or transgressing a convention we can exercise control over it, this control must be indirect; that is, conformity/transgression must trigger some further thing which in turn can either sustain or abolish the convention. But is this the case? Cohen points out that both conformity and transgression can have a pressurising effect on others, and he seems to think that this pressure is what enables us to influence conventions. By conforming to or transgressing a convention, I can invoke others around me to act likewise; but since conventions exist only insofar as people act in accordance with them, if sufficiently many transgress a given convention, then it will cease to be one.

All of these points may be true. But they do not establish what Cohen needs: that by conforming to, or transgressing, a convention, an individual can bring something about which will cause that convention to persist or cease to be. Granted, certain others might be disposed, or have reason, to imitate them.

But it will be very rare that an individual can influence enough people to ensure either the convention's future survival, or its end. For one thing, it is rare that they will be able to publicise their act to sufficiently many others (though developments in communication technology are making this easier). But even if they could, the pressure to conform with one's peers is only one among several factors which can influence people's behaviour. For instance, self-interest might move people to act in accordance with a prevailing convention, even if others around them transgress it; a husband might continue to do less housework than his wife because the arrangement suits him, even if some of his male friends have opted to take on more. Indeed, even if a conventional arrangement is detrimental to some individuals, they may still be moved to comply with it: perhaps by the short-term benefits of doing so, or perhaps by 'status quo bias', an aversion to change. Jean-Claude Kaufmann (1998), for instance, argues that young, heterosexual couples can slide into unequal, gendered divisions of housework as a strategy for maintaining relationship stability: even if the woman is aware that doing so will be detrimental to her in the long run.

To be clear, this is not to say that it is impossible to possess indirect control over informal social conventions. Notably, one mechanism through which such control can be exercised is the law. Legislators can wield indirect control over informal conventions in a fairly straightforward way: if they make it illegal to act in accordance with a convention, and individuals tend to comply with the law, then that convention will be unlikely to garner enough conformity to ensure its continued survival. But the law can also have more subtle effects on conventional behaviour. Legislative changes can prompt individuals not only to re-evaluate whether they have reason all things considered to engage in a certain practice, but also to reassess their independent reasons for participating in it.

So, if all of this is correct, then the conventionalist reading of the basic structure will not imply that the principles of justice make requirements on individuals across a broad range of quotidian circumstances. Given the conventionalist reading, applying these principles to the basic structure will

imply that they require individuals to shape informal social conventions so that the resulting distribution of rights, duties, and benefits is just. But the range of circumstances in which an individual will be capable of doing this will be much less broad than Cohen seems to have imagined; no one can control informal conventions directly, and it is rare that a person will be in a position to exercise indirect control over them. As such, the second horn of Cohen's dilemma poses no threat to my reading of the basic structure restriction; even if the conventionalist reading of the basic structure is correct, applying the principles of justice to it will still generate a scope distinction between those principles and the principles for individuals.[25]

[25] Perhaps Cohen's thought is not that each of us *individually* is able to control prevailing conventions, but rather that all of us *collectively* are able to do so. Given this, the thought might continue, the principles of justice require us not to conform to conventions which sustain unjust distributions of rights, duties, and benefits: since each act of conformity is one among many which together cause such conventions to persist. I will address this kind of thought in Chapters 4 and 5. The overall argument of these two chapters is that an individual does not have a reason to refrain from an action, purely in virtue of the fact that it is one among several which collectively cause harm.

2

Equal Opportunity

The preceding chapter focussed on a general feature of Rawls' principles of justice: that they are distinguished from other moral principles by the fact that they apply primarily to the basic structure of society. This chapter concentrates on one of these three principles of justice which received comparatively little attention in Chapter 1: the principle of fair equality of opportunity. In Section 1, I expound a number of difficulties which face this principle. However, I suggest that these can be avoided if we adopt a theory of equal opportunity recently proposed by T. M. Scanlon. This theory has not attracted the level of attention which I believe it deserves. Section 2 aims to redress this: I offer a detailed analysis of Scanlon's account, and of how it resolves the difficulties facing Rawls' conception of equal opportunity. Section 3 considers the requirements which a Scanlonian principle of equal opportunity would make of individuals, assuming that it is subject to the basic structure restriction (as interpreted in Chapter 1). More specifically, I consider whether the requirements of such a principle would conflict with a plausible principle of parental partiality. I consider three areas in which conflicts between equal opportunity and parental partiality have been thought to arise: the acquisition of unfair advantages, school choice, and the justification of the family. In each case, I argue that such conflicts dissolve if equal opportunity is conceived in line with Scanlon's theory.

1. Rawls on equal opportunity

Rawls' conception of equal opportunity has two components, which are often referred to as its *formal* and *substantive* elements. The formal element comprises of what Rawls refers to as "careers open to talents" (Rawls 1999, p. 57). This requires that "all have at least the same legal rights of access to all advantaged social positions" (Rawls 1999, p. 62). Rawls supplements this formal requirement with the more substantive requirements of the principle

of fair equality of opportunity, which requires that "positions are to be not only open in the formal sense, but that all should have a fair chance to attain them" (Rawls 1999, p. 63). This broad requirement is fleshed out as follows:

> More specifically, assuming that there is a distribution of natural assets, those who are at the same level of talent and ability, and have the same willingness to use them, should have the same prospects of success regardless of their initial place in the social system. In all sectors of society there should be roughly equal prospects of culture and achievement for everyone similarly motivated and endowed. The expectations of those with the same abilities and aspirations should not be affected by their social class (Rawls 1999, p. 63)

Rawls' principle captures a conception of equal opportunity which is familiar and intuitive. Taken together, its formal and substantive components demand (a) that all individuals have the same legal rights to be considered for any position for which they choose to apply, and (b) that equally talented and motivated individuals have the same prospects of attaining positions of advantage, regardless of their social background. In essence, these requirements map roughly onto a commitment to non-discrimination, and to eliminating the influence of social class on people's life prospects.

However, as it stands, there are problems with the principle. Consider first the formal requirement that all have the same legal rights of access to all positions. Presumably, this would prohibit an employer from choosing not to consider someone for a position because of the colour of their skin, or their sexual orientation, or their gender etc. But it would not prohibit them from, say, rejecting a candidate for the position of French teacher on the grounds that the candidate cannot speak French. The formal requirement would, presumably, also proscribe an employer from offering a position to a candidate because they are an old friend of the candidate's father, or because the candidate is willing to offer sexual favours etc. But it would not prevent them hiring a candidate on the basis of certain other qualities, such as

proficiency in a foreign language, or the possession certain qualifications etc. That is, formal equality of opportunity will constrain the reasons for which an employer can permissibly hire or reject a candidate. A certain class of considerations will be considered 'legitimate' reasons for hiring/rejecting, whilst others will be considered 'illegitimate' or 'arbitrary'.

All of this should be obvious enough. But it invites a question: What distinguishes the 'legitimate' reasons for hiring/rejecting from the 'illegitimate' ones? Why is it that it is permissible to reject a candidate for the position of French teacher on the grounds that they cannot speak French, but not on the grounds that they are a woman? Why is an employer permitted to hire someone for the reason that they have good foreign language skills, but not for the reason that they are willing to offer sexual favours? In short, what we need is a principled way of drawing the distinction between the legitimate forms of discrimination which, presumably, the formal element of equal opportunity permits, and the arbitrary forms which it prohibits. Rawls' brief discussion of careers open to talents does not give us this.[26]

But the problems don't stop with formal equality of opportunity. Further difficulties arise when we scrutinise the conception's substantive element. According to Rawls, this requires that equally talented and motivated individuals have equal prospects of attaining an advantageous position, regardless of their social background. That is, on the one hand, people's prospects of attaining a given position are permitted to differ according to differences in their levels of ability and motivation. But, on the other, they are not permitted to differ according to differences in their social backgrounds. On the face of it, this might seem like precisely what we want from a principle of equal opportunity; we want it to ensure that talented individuals who strive to attain positions of advantage will be neither helped nor hindered by the contingencies of social class. However, on closer inspection, it seems that these two aims of substantive equal opportunity could come into conflict.

[26] Richard Arneson (1999, pp. 78-80) has argued that the Rawlsian notion of careers open to talents may permit discriminatory practices which intuitively seem objectionable. I will consider a worry similar to Arneson's in Section 2.2.

To illustrate, consider on what grounds we might say that two individuals are equally talented, or of equal ability. One way in which to determine this might be to set some sort of examination for both of the individuals to take, and to compare their scores. If the individuals scored differently, then substantive equality of opportunity would permit their prospects of attaining certain advantageous positions to differ accordingly; that is, it would permit the higher-scoring candidate to have better prospects than the lower-scoring candidate. However, now suppose that people's test scores turn out to be strongly correlated with social class: higher-scoring individuals consistently come from upper- and middle-class backgrounds, whilst lower-scoring candidates are consistently drawn from working-class families. Given this, the two aims of Rawls' principle of substantive equality of opportunity will conflict with one another. If we treat test scores as indicative of ability, then the principle will permit people's prospects to differ when they have different test scores. But if how well people score is strongly correlated with their social class, then the principle will also prohibit people's prospects to differ in this way: since it does not permit prospects to differ according to social background.[27]

The same problem arises when we consider motivation. Rawls' principle of substantive opportunity permits the prospects of equally talented individuals to differ if they are not equally willing to use those their talents. Yet, just as the markers of ability might be correlated with social background, a person's willingness to make use of their abilities can also vary according to their family background. It seems not unreasonable to expect that the most motivated individuals will be drawn from the higher social strata, whilst those least willing to apply themselves will be drawn from the least well-off families. If so, then Rawls' conception of substantive opportunity will engender a further tension; it will be impossible both to allow people's prospects to vary according to their willingness to make use of their abilities, and to prevent them from differing according to social background.

[27] This point is similar to one raised by Bernard Williams (1962, pp. 125-128).

Rawls was fully aware of these tensions.[28] Indeed, the difficulty of insulating ability and motivation from the influence of social background is one of the reasons he gives for supplementing the principle of fair equality of opportunity with the difference principle (Rawls 1999, p. 64). Nevertheless, it seems worthwhile to ask whether the former principle might be developed in a way which could mitigate those tensions.

2. Scanlon's theory

T. M. Scanlon (2018, chaps. 4, 5) offers an account of equal opportunity which develops the conception found in Rawls. He asks us to imagine that someone objects of being worse off than other people. A satisfactory response to that objection, Scanlon says, must involve three claims:

> 1. *Institutional Justification*: It is justified to have an institution that generates inequalities of this kind.
> 2. *Procedural Fairness*: The process through which it came about that others received this advantage while the person who is complaining did not was procedurally fair.
> 3. *Substantive Opportunity*: There is no wrong involved in the fact that the complainant did not have the necessary qualifications or other means to do better in this process (Scanlon 2018, p. 41)

According to Scanlon, the key to understanding the ideal of equal opportunity lies in coming to understand each of these claims, and the relations between them.[29]

---

[28] "The extent to which natural capacities develop and reach fruition is affected by all kinds of social conditions and class attitudes. Even the willingness to make an effort, to try, and so to be deserving in the ordinary sense is itself dependent upon happy family and social circumstances" (Rawls 1999, p. 64).

[29] Scanlon assumes that complaints grounded in the notion of equal opportunity arise in cases in which one person is worse off than another. But one might question this assumption. Suppose that you and I have equally well-paid jobs. However, because of the different colour of our skin, we are not permitted to occupy one another's roles. Couldn't both of us legitimately complain that our opportunities are not equal? In fact, Scanlon maintains a neutral position on this question. He distinguishes between two types of rationale for a

The similarities between Scanlon's conception of equal opportunity and Rawls' are more or less obvious from the start. The second and third parts of the three-part justification of inequality, procedural fairness and substantive opportunity, roughly correspond respectively to the formal and substantive elements of Rawls' conception. And the first part, institutional justification, marks the place which in Rawls' theory is occupied by the difference principle: the principle which regulates the distribution of income and wealth produced collectively by society's major institutions. However, each element of Scanlon's account differs in important ways from Rawls' treatment. As such, it will be instructive to consider each in turn.

2.1. Institutional justification

Let's begin with the first element: institutional justification. Scanlon considers a number of ways in which one might try to justify some inequality-generating institution. However, he focusses on one form of justification in particular, according to which an inequality-generating institution is justified by appeal to the benefits it produces (Scanlon 2018, p. 41). For instance, suppose that a health service pays its doctors and nurses a higher wage than that which certain other institutions pay their employees. The administrators at the health service might justify this wage differential by claiming that by paying doctors and nurses more, they can attract the best candidates for these roles. Moreover, everyone benefits from this; everyone gets better doctors and nurses.

There are a number of possible variations on this form of justification. The foregoing justification of a wage differential between healthcare workers and

---

principle of equal opportunity: the *just inequality rationale* and the *self-realisation rationale* (Scanlon 2018, pp. 56-57). According to the former type of rationale, the requirements of equal opportunity need only apply to those positions to which greater rewards and privileges are attached. However, rationales of the self-realisation type imply that these requirements apply to all positions, regardless of whether special rewards are attached to them. Scanlon focusses on the narrower sort of principle implied by the just inequality rationale, and I will follow him in this section. However, he does not deny that it might be possible to justify the more expansive sort of principle implied by the self-realisation rationale. In Section 3.3.3, I will consider a potential rationale for a principle of equal opportunity which is of the self-realisation form.

other professionals claimed that *everyone* benefits from that inequality. But one could propose to justify certain wage differentials by appealing only to the benefits it provides for a specific subset of individuals. Rawls' difference principle functions in this way. This principle states that it is justifiable for an institution to generate inequalities in wealth only if these inequalities benefit the worst off. As such, wage differentials are to be justified by reference to the benefits they generate for the worst-off class of individuals.

2.2. Procedural fairness

The second level of Scanlon's three-part justification of inequality, procedural fairness, is to be understood in light of the first level. Suppose that the justification for doctors and nurses being paid more than certain other workers is that this wage disparity is to the greatest benefit of the least well off. Given this, Scanlon (2018, p. 42) claims that the procedure by which doctors and nurses are selected would be fair if and only if that procedure did in fact procure those benefits for the worst off. To see the implications of this claim, consider the details of the proffered justification. This justification alleges that paying higher wages for healthcare workers benefits the worst off because it attracts individuals who are best suited to those roles, and having such individuals fill those roles is beneficial to the worst off. Given this, the health service's selection process will be fair, according to Scanlon, if and only if it selects individuals on the basis of characteristics which do in fact best suit them to being doctors or nurses.

Scanlon (2018, p. 44) refers to this as an "institution-dependent" notion of talent: selection procedures are fair just if they select on the basis of talent, and what counts as a talent depends upon how the inequalities which the institution in question generates are justified. As such, the capacities on the basis of which an institution can permissibly select will depend upon the goal of that institution (to provide healthcare, for instance). Perhaps less obviously, it will also depend on the organisational possibilities open to it. For instance, given one way of organising work within a certain institution, an individual might need a certain level of physical strength in order to carry out a particular job. But if mechanical aids were employed, then this would no longer be a

requirement. Does this mean that this institution could permissibly select on the basis of physical strength, so long as it refrains from introducing mechanical aids? Not necessarily. If this institution could implement such aids without compromising the goal which justifies its wage disparities, then physical strength would not count as a 'talent' on the basis of which candidates could be fairly selected: it would be possible for workers to produce the justifying benefits without being physically strong. As such, Scanlon claims that it would be impermissible for the institution to continue to select on the basis of physical strength; procedural fairness requires this institution to organise itself in a way which does not require physical strength of its employees (Scanlon 2018, p. 46).

Scanlon's institution-dependent account of procedural fairness develops the formal element of Rawls' conception of equal opportunity in a way which resolves the difficulty noted in the preceding section. That is, Scanlon offers a principled way of distinguishing between legitimate and illegitimate forms of discrimination: it is fair to select/reject a candidate for a given institutional position only on the grounds that they possess/lack the qualities necessary for them to be able produce the benefits which justify the inequalities produced by that institution.

At this point, one might raise an objection. It seems that wrongful discrimination could occur without the form of procedural unfairness which Scanlon describes. For instance, to adapt an example of Seana Shiffrin's (2004a, p. 1647), we would be hard pressed to find any procedural unfairness in the practice of directing African Americans to sit at the back of buses. And yet this practice seems archetypal of wrongful discrimination. However, it is important to note that Scanlon is not offering a general account of what makes wrongful discrimination wrong. Rather, he simply isolating one type of wrong which could be committed when employers arbitrarily discriminate between candidates (Scanlon 2018, pp. 43-44). As such, he can allow that there are many other things to object to about wrongful discrimination besides procedural unfairness.

However, there are cases of a certain structure which do seem to pose a problem for Scanlon. These are cases of what John Gardner (2018a, p. 69) has called "inherited discrimination", in which the fact that one individual discriminates against another for illegitimate reasons seems to give some third individual legitimate reasons for discriminating in the same way. For example, suppose that the locals of a particular region are generally very xenophobic. Indeed, xenophobia is so rife that most of the locals will refuse to receive medical treatment from a foreign doctor. The locals' refusal, on xenophobic grounds, to be treated by foreign doctors is, let's assume, an instance of wrongful discrimination. However, given Scanlon's account of procedural fairness, it seems to provide the healthcare service with a legitimate reason not to hire foreign doctors. If most of the locals would refuse to be treated by a foreigner, then a foreign doctor working in this region would be unable to provide care to many of the people who need it. But if so, then hiring foreign doctors would leave the healthcare service unable to provide the benefit (healthcare) which justifies its wage disparities. To put the point in potentially vexing terms, not being foreign would, in this case, constitute an institution-dependent talent.[30]

Why does this example pose a problem for Scanlon? It seems that foreign individuals who have the same capacities which would qualify a native individual for a position in the healthcare service can legitimately complain about their situation. Moreover, their complaint is not merely that they are subject to xenophobia. This would be their only complaint if they were hired as doctors but were continually rebuffed by the xenophobic locals. However, if the healthcare service refuses to hire foreign doctors, then these individuals seem to have a further complaint: one which seems to be based in the ideal of equal opportunity. That is, it seems that qualified foreign individuals would be justified in complaining that they do not enjoy the same opportunities as their native counterparts. However, if Scanlon's theory implies that the

[30] A similar example is discussed by Arneson (1999, pp. 79-80) in relation to Rawls' conception of equal opportunity.

healthcare service's hiring procedure is fair, then how can he make sense of this complaint?

Scanlon, I think, has the resources to respond to this objection. However, in order to see this, we will need to move on to address the final element of his conception of equal opportunity.

2.3. Substantive opportunity

The third and final component of Scanlon's three-part justification of inequality, substantive opportunity, builds on the preceding two. At base, it stipulates that individuals must not be wrongfully blocked from attaining institution-dependent talents; that is, individuals must not be wrongfully prevented from acquiring those qualities on the basis of which employers may permissibly select/reject candidates.

At first glance, this principle seems open to two objections. First, it might seem redundant. Why do we need a moral principle to tell us not to wrongfully prevent people from attaining institution-dependent talents? Proposing such a principle seems like proposing a moral principle which tells us to commit no wrong. The fact that an action is wrong already implies that we ought not to do it. A principle which simply repeats this implication would be redundant. Second, Scanlon's principle of substantive opportunity might seem too general to be informative. An action which prevents others from developing institution-dependent talents could, it seems, be wrong for a variety of different reasons. As such, one might hope that an account of substantive equality of opportunity would isolate some of these wrongs, and spell out their implications for the organisation of the basic structure of society.

Let's take the first objection first. Scanlon's principle of substantive opportunity would be redundant if it simply directed us to refrain from wrongfully blocking others from developing institution-dependent talents. But to read the principle as such is to misinterpret it. It is not intended as a stand-alone moral principle, but rather as part of a reply to someone who

objects of being worse others than others. It stipulates that if that person was wrongfully prevented from developing certain qualities which would have increased their chances of being fairly selected for advantageous positions, then their complaint is legitimate.

Interpreted in this way, substantive opportunity is not redundant. To see this, consider how it might be invoked to make sense of the complaint which it seems could be raised by individuals in the case of inherited discrimination discussed above. Recall that these individuals are foreign nationals who, if they did not happen to be foreign, would qualify for positions in the healthcare service. However, this institution has decided not to hire foreign healthcare workers on the grounds that, if it did, it would be unable to deliver the benefits which justify its wage disparities. Earlier, I said that these individuals could, it seems, legitimately complain that their opportunities are worse than other people's. On Scanlon's account, we cannot make sense of this complaint in terms of procedural fairness. But we can, I suggest, make sense of it in terms of substantive opportunity. The fact that nationality counts as an institution-dependent talent is explicable by reference to wrongful behaviour: namely, the xenophobic behaviour of the locals. As such, we can describe the complainants as wrongfully prevented from progressing in the fair selection procedure employed by the healthcare service; if it weren't for the locals' xenophobia, they would qualify for selection. This complaint goes beyond merely being subject to xenophobia. It is a legitimate complaint about inequality, a complaint that these individuals are worse off than others.

The second worry is better taken than the first. Does Scanlon offer anything further which might help us individuate some of the principal ways in which people's opportunities to develop institution-dependent talents can be wrongfully blocked? In fact, he does. Scanlon (2018, p. 58) appeals to Rawls' principle of fair equality of opportunity, and specifically his stipulation that "those who are at the same level of talent and ability, and have the same willingness to use them, should have the same prospects of success regardless of their initial place in the social system" (Rawls 1999, p. 63). Given this, if two equally talented and motivated individuals from different social

backgrounds end up with different prospects of being fairly selected for positions of advantage, then the worse off of the two will have been wrongfully blocked from acquiring institution-dependent talents.

Initially, this claim might seem odd. If the two individuals are equally talented and motivated, then how could they end up with different prospects of being fairly selected? A fair selection procedure is one which selects on the basis of institution-dependent talents. But we've already assumed that both of these individuals are equally talented and equally willing to use their talents. As such, it would seem to follow that they must have equal prospects of being selected by a fair selection procedure. But if so, then Scanlon will get Rawls' principle for free. That is, this requirement will not be able to elaborate Scanlon's notion of substantive opportunity, since it is already implied by the prior notion of procedural fairness.

However, this is too fast. We need not interpret 'equally talented' to mean 'possessing the same institution-dependent talents'. Instead, we can take it mean that the two individuals in question *would* develop the same institution-dependent talents, if they were provided with the same quality of education and training (Scanlon 2018, pp. 59-60). Interpreted as such, if two equally talented and motivated individuals were to end up with different prospects of fair selection, then this would imply that they had not had access to the same level of education and training. Thus, to say that the worse off individual in this instance would have been treated wrongfully would amount to saying that everyone is owed a certain level of education: namely, an equal one.

This way of understanding 'equally talented' provides us with a requirement which goes beyond those of procedural fairness. The latter notion requires employers to select candidates on the basis of institution-dependent talents. But it does not require anyone to provide a level of education which enables individuals to acquire those talents. This requirement, I am suggesting, is made by the substantive element of equal opportunity.

However, more than this, the current proposal can also help to resolve one of the tensions in Rawls' conception of equal opportunity. In the preceding section, I pointed out that if the markers of ability can be influenced by a person's social background, then it will be contradictory to permit people's prospects to differ according to talent whilst prohibiting them from differing according to social background. But on the proposed reading of 'equally talented', the markers of talent are already insulated from the influence of social background. Suppose that a child from a wealthy background is more likely to develop institution-dependent talents than a poorer child because their family can afford to pay for a higher quality of education. Would a principle of substantive opportunity imply that it is permissible for the prospects of these two individuals to differ? If we understand 'equally talented' in the way being proposed, then it will not. The wealthy child will not necessarily count as more talented than the poorer child, since this comparison can be made only under the assumption that both children receive an equal level of education. So, if wealthy parents pay for their children to be better educated than poorer children, then we cannot use the talents which these children actually develop to determine whether it is permissible for their prospects to differ; the relevant judgement to make is whether or not these children *would* have developed the same talents given an equal standard of education.

So, understood as I am proposing, substantive opportunity requires that all members of society be provided with an equal standard of education and training; the quality of the educational opportunities to which one has access must not depend on one's social background.[31] But notice that this is consistent with claiming that it is permissible for no one to be provided with any level of education at all. That is, if no one is educated, then everyone is

[31] In fact, this requirement is pre-empted by Rawls: "Chances to acquire cultural knowledge and skills should not depend upon one's class position, and so the school system, whether public or private, should be designed to even out class barriers" (Rawls 1999, p. 63). As to whether this means that Scanlon's account simply offers a more accurate interpretation of Rawls' conception of equal opportunity, or whether it provides Rawls with the necessary tools for saying what he wanted to say, I will remain agnostic.

educated equally well. Does substantive opportunity require anything more than an equally bad, or non-existent, education for all?

According to Scanlon (2018, pp. 59-60), we should judge two individuals to be equally talented just if they would develop comparable institution-dependent talents if they were provided with the standard of education which the wealthiest bracket of society can afford for their children. This would imply that substantive opportunity requires everyone to be provided with that very high standard of education. However, it's not clear why we should follow Scanlon here. So long as the standard of education is equal across society, the influence of social class on people's prospects of being fairly selected for positions of advantage will be assuaged; and it will be no more assuaged if we set that equal standard at the highest level available, than if we set it at some lower level.

Nevertheless, I suggest that Scanlon's theory does have the resources to avoid the conclusion that equality of opportunity permits a very low standard of education. One possible strategy would be to refer back to the first level of his three-part conception, institutional justification. Here, recall, Scanlon claimed that we can justify the inequalities in income and wealth generated by a particular institution by appealing to the benefits which those wage disparities produce. One way in which a wage disparity can benefit others is by attracting talented (in the institution-dependent sense) individuals to work in the higher-paid roles. However, if the standard of education and training is so poor that no one develops the requisite talents, then the institution in question will be unable to procure such benefits. So, if we have reason to want institutions to procure the benefits which justify having wage disparities, then we have reason to want the societal standard of education to be good enough to enable people to acquire institution-dependent talents.

A further strategy can be inferred from the way in which Scanlon proposes to resolve the second tension in Rawls' conception of equal opportunity which I pointed out in the preceding section. Recall that this tension arises because it seems that a person's willingness to use their abilities can be influenced by

their social background. But if so, then it will be contradictory to permit people's prospects to differ according to differences in motivation, but not to permit them to differ according to social background.

Scanlon (2018, p. 60) points out that there is an ambiguity in Rawls' claim that people's prospects can permissibly differ according to their willingness to use their talents. On the one hand, we could interpret this claim as saying that it is always permissible for one person to be worse off than another if they have failed to make the efforts which that other person has. On this reading, people's wealth and prospects can permissibly differ according to the actual amount of effort they put into acquiring institution-dependent talents and competing for positions of advantage. But, on the other hand, we could also interpret Rawls' claim as saying that it can be permissible for people to be better or worse off than one another according to the choices which they make.

Initially, it might be difficult to distinguish these two readings. Both seem to be saying that we should allow people's prospects to differ depending on whether or not they choose to take advantage of the opportunities on offer. The difference, however, lies in *which* choices we think should determine people's relative prospects. On the first reading, any difference in the effort people put into attaining positions of advantage, whether this be in acquiring institution-dependent talents or in competing with others in fair selection procedures, is permitted to lead to inequalities in wealth, or in future prospects for acquiring wealth. This reading is liable to lead to the tension just mentioned. People's willingness to make an effort can be influenced by their social background; and so permitting people's prospects to differ accordingly can conflict with the aim of preventing them from differing according to social background.

On the second reading, however, not all choices need be permitted to lead to inequalities. That is, we can think that people's relative prosperity should sometimes depend on the choices they make, without thinking that *all* choices should have such influence. Scanlon (1998, pp. 251-256; 2018, p. 61) points

out that we can have a variety of reasons for wanting what happens to us to depend on our choices. Some may be instrumental. For example, I may have reasons for wanting the dish a waiter serves me to correspond to that which I chose from the menu, because it's more likely that I'll enjoy eating that one, or because I won't swell up when it touches my skin (Scanlon 1998, pp. 251-252). But some of these reasons may be of a non-instrumental nature. For instance, I might have reasons for wanting my career to be one which I have chosen, even if it turns out that I would be better at a job picked out for me by someone else. Perhaps this is because I have reason to want my career to reflect my own particular interests and capacities (Scanlon 1998, p. 253). Or it might be that simply having the opportunity to make such a choice is itself valuable; perhaps this opportunity expresses that others think I am capable of making important decisions (Scanlon 1998, p. 253), or perhaps having such opportunities is necessary for people to lead autonomous lives (Raz 1986, chaps. 14, 15).

This second reading offers a way of resolving the apparent tension in Rawls' claim. We've seen that some of the choices which affect people's prospects can be influenced by their social background. For example, an individual's decision to drop out of education early might be influenced by the fact that formal education was not valued highly within their family; and this decision may go on to have a large impact on their relative prosperity. However, if it can be shown that we do not have reason to want our prospects to depend on such choices, then the claim that people's prospects can permissibly differ according to their choices will not lead us into conflict. This claim will not contradict the claim that it is impermissible for people's prospects to differ according to their social background, because the choices which should be permitted to lead to inequalities will not include those which are influenced by differences in social background.

But can it be shown that we have no reason to want our prospects to be determined by choices which are influenced by our social background? Scanlon offers a possible argument for this claim. He submits that we have reason to want our prospects to depend on our choices only when we have

sufficiently good knowledge of the options open to us (Scanlon 2018, p. 62). I might not have reason to want the dish I am served to correspond to that which I ordered if I am unacquainted with the items on the menu. In this case, it might be better for a friend, or the chef, to choose for me. Similarly, I might not have reason to want my career to be one I have chosen, if am not properly informed of the alternatives. This choice would not reflect my interests, if there were other options of which I was unaware which would better suit them. Nor would the opportunity to make this choice express a judgement that I am capable of making such decisions, or enable me to take authorship over my life, if information relevant to the making of it was withheld. However, when a person's prospect-affecting choices are influenced by their social background, they are often made under conditions of partial information. For instance, when a person decides to drop out of education early because of their family's negative attitudes towards formal education, their decision is made without sufficient information of the full range of options open to them. Specifically, the supposition that formal education lacks value dominates the way in which they frame the choice-situation facing them. Had they had access to a richer framing of their options, they might have chosen differently.

So, Scanlon's arguments provide a way of resolving the tension between permitting people's prospects to differ according to their choices, but not according to their social background. If the choices which are influenced by social background are made without sufficient knowledge of the relevant options, then we need not permit people's prospects to be determined by them. But moreover, I suggest that they also provide a way of avoiding the conclusion that equality of opportunity permits a very low standard of equal education. Ideally, we would want people's choices to have value; that is, we would want people's epistemic conditions to be good enough that they can have reasons, of instrumental and non-instrumental kinds, to want their prospects to be determined by the choices which they make. But if this is correct, then it follows that people ought to be offered comprehensive information about the variety of opportunities on offer in their society. Offering such information would involve providing an education which trains

individuals in a wide range of disciplines, and which informs them of the various valuable ways in which they might structure their lives. Whilst such an education need not be the very best which money can buy, it would be a lot better than the worst which could be offered.

3. Parental partiality

The preceding two sections developed and analysed a particular way of construing the ideal of equal opportunity. On this conception, equality of opportunity involves a formal element which prohibits employers from discriminating between candidates on arbitrary grounds. It also contains a substantive element, which requires that everyone have a fair chance of acquiring those attributes on the basis of which employers may legitimately select. Achieving this involves providing an equal standard of education for all, as well as ensuring that this education informs individuals of the diversity of valuable ways of living.

This conception of equal opportunity coheres with Rawls' claim, defended in Chapter 1, that the principles of justice apply primarily to the basic structure of society. I argued that this claim should be understood as saying that the principles of justice require individuals to ensure that the distribution of fundamental rights, duties, and benefits determined collectively by society's major institutions cohere to the specifications of those principles. Applied to the conception of equal opportunity defended by Scanlon, this implies that individuals are required to ensure: (a) that institutions do not generate unjustifiable socio-economic inequalities; (b) that employers within those institutions discriminate between candidates only on the basis of institution-dependent talents; and (c) that the major institutions do not, either singly or collectively, wrongfully block individuals from attaining institution-dependent talents.

In this section, my aim will be to assess the extent to which these individual requirements conflict with those of a further normative principle: a principle of *parental partiality*. Such a principle states that parents have reasons to treat their own children in certain preferential ways. This idea is deeply intuitive.

Parents are typically motivated to do things for their own children which they are not motivated to do for the children of others; they look out for their children's welfare in ways that they do not for strangers'. Moreover, we do not view this as irrational, or morally abhorrent. Showing partiality for one's own children, we think, is part of what it is to be a good parent. Of course, there are limits to the partiality we think parents have reason to show towards their children. A parent should not, for instance, kill a stranger's child in order to better their own child's chances of winning a competition. Nonetheless, we do think that parents have reason to show their children certain restricted forms of preferential treatment.[32]

A principle of parental partiality seems apt to conflict with the requirements of a principle of equal opportunity, as construed in preceding sections. The latter principle stipulates that people's prospects of attaining positions of advantage should not be differentially influenced by their family backgrounds. As such, it seems natural to think that it will imply restrictions on the advantages which parents can procure for their children, so as to ensure equality in the influence which families have over individuals' prospects. However, intuitively, it does not seem that a parent will overstep the boundaries of legitimate partiality merely by providing her own child with benefits which are not offered by other parents; a plausible principle of parental partiality must, it seems, allow that parents can have reasons to benefit their children, even if there are other children who do not receive those same benefits from their own parents. But if so, then it seems that conflicts are bound to arise between the permissions implied by the principles of equal

[32] Niko Kolodny (2010) outlines two possible accounts of partiality principles in general. One might adopt a normative theory on which all fundamental normative principles are impartial: consequentialism, for instance. On such a view, all valid partiality principles must be derivable from the fundamental, impartial normative principles. For example, one might propose that the principle that one has reason to show parental partiality towards one's own children can be derived from the more fundamental principle that aggregate wellbeing ought to be maximised, since parents better promote the wellbeing of their own children. Alternatively, one might propose that there are fundamental partiality principles. On a view of this sort, the valid partiality principles need not be shown to follow from further impartial normative principles. Kolodny himself defends a view of the latter type. Here, however, I am not concerned to adopt one or the other of these views. I merely aim to highlight the pre-theoretical plausibility of a principle of parental partiality. For further discussion of the grounds for, and limits to, such a principle, see (Brighouse and Swift 2009).

opportunity and parental partiality: equal opportunity seems to prohibit parents from doing things for their children which will give them an advantage over others, whilst parental partiality seems to imply that they can have reason to do just that.

The prospect of such a conflict should be worrying. On the one hand, equality of opportunity is a part of an attractive ideal of social justice. But on the other, the claim that we have reason to show preferential treatment towards our own children is deeply intuitive. As such, it seems that neither is something we should be willing to give up on.[33]

The following three subsections address three areas in which a conflict between equal opportunity and parental partiality seems to arise: situations in which parents can help their children to acquire attributes which will improve their prospects of success in an unfair competition; situations in which parents can purchase educational advantages for their children; and the justification of the family as the primary source of childcare and socialisation. In each case, I will argue that, contrary to appearance, the requirements of equality opportunity and parental partiality are not in tension. Moreover, this is not because the limits of permissible parental partiality are more restricted than they might at first appear, but rather because equal opportunity requires less of parents than has sometimes been thought.[34]

3.1. Acquiring unfair advantages

Suppose that the number of highly advantageous institutional positions in a society is relatively small. Indeed, there are much fewer such positions than individuals who could competently occupy them: that is, who could provide the social benefits which justify the higher remuneration attached to those

[33] This thought bears some similarity to Thomas Nagel's (1991, chap. 2) contrast between the personal and impersonal standpoints. Nagel argues that we grasp distinct values by alternating between these two standpoints, and that an acceptable theory of political morality must find a way to consolidate those values.

[34] Harry Brighouse and Adam Swift (2009), for instance, argue that conflicts between equality of opportunity and legitimate parental partiality are less common than is often thought. However, for them, this is largely because the values realised by the parent-child relationship commonly fail to justify parents in acting in ways which would undermine equal opportunity.

roles. In order to whittle down this large pool of competent candidates, institutions introduce selection processes which discriminate between candidates on the basis of qualities which are strictly superfluous to the roles on offer. For instance, they might require candidates to possess postgraduate qualifications, even though someone with only an undergraduate education could perform the tasks in question sufficiently well.

Given Scanlon's account of procedural fairness, these selection processes are unfair: not all of the qualities they require of candidates constitute institution-dependent talents. As such, a principle of equal opportunity which encompasses this requirement would obligate the employers in question to revise their selection procedures, in order to make them fair. However, Scanlon suggests that it is not only employers who would be subject to the requirements of such a principle. When employers unfairly select candidates for advantageous positions on the basis of arbitrary qualities, wealthy parents, he claims, are required not to procure those qualities for their children. That is, if employers unfairly discriminate between applicants on the basis of whether or not they possess a master's degree, or have experience working in unpaid internships, or attended some prestigious summer school etc., then parents are required not to use their wealth in order to acquire these things for their children. Parents who do use their wealth in these ways can be accused of "gaming the system"; they take advantage of extant unfairness within the job market, and as such interfere with the achievement of equality of opportunity (Scanlon 2018, p. 69, fn. 30).

If procedural fairness requires this of parents, then it would seem to conflict in one obvious way with a principle of parental partiality. Scanlon is suggesting that procedural fairness requires parents not to procure for their children benefits on the basis of which employers unfairly select for positions of advantage. However, it seems not unreasonable to think that a plausible principle of parental partiality would allow that parents can have reason to benefit their own children in these ways. There are two possible grounds for thinking this. First, one might think that parents have reason to want their children to fare as well as possible in any competition for positions of

advantage. If so, then they would have reason to enable their children to acquire those qualities on the basis of which candidates for such positions are selected, even if some of those selection procedures are unfair. However, one need not subscribe to this strong view in order to think that parents have reasons of partiality to provide for their children in this way. The qualities which employers unfairly use to discriminate between candidates may be valuable in ways beyond their instrumental value for attaining positions of advantage. As such, parents need not have reason to want their children to have the best possible prospects of winning an unfair competition, in order to have reason to procure these benefits for their children.

So, on Scanlon's view, procedural fairness and parental partiality can come into conflict, at least under conditions of procedural unfairness. He is not explicit about which, if either, consideration ought to take precedence in cases of conflict.[35] But however we answer this question, we face an uncomfortable conclusion: under conditions of injustice, a commitment to achieving social justice is not always compatible with parental partiality.

One immediate question to ask is whether or not Scanlon's claim is consistent with the Rawlsian view that a principle of equal opportunity, as a principle of social justice, must apply primarily to the basic structure. More specifically, one might wonder why the requirements of procedural fairness do not apply exclusively to employers and legislators. They are the ones who are in a position to alter institutions' selection procedures: the employers by revising the institutional rules directly, the legislators by requiring, or incentivising, employers to do so. Why, then, do these requirements extend to parents, who may lack any control over institutional selection procedures?

[35] He writes: "But if procedural fairness is not achieved, then providing these benefits [i.e. those on the basis of which employers unfairly discriminate] for one's children, however irresistible it may be, is a way of gaming the system" (Scanlon 2018, p. 69 fn. 30). But this doesn't necessarily settle the question of what parents ought to do. Should they refrain from gaming the system? Or is it sometimes permissible for them to succumb to the temptation to provide these benefits for their children, even if doing so interferes with procedural fairness?

Scanlon can meet this worry, I suggest, by phrasing the individual requirements of procedural fairness in the way presented at the outset of this section. I said that this component of equal opportunity requires individuals to ensure that employers discriminate only on the basis of institution-dependent talents. Phrased as such, procedural fairness is consistent with the basic structure restriction, since it directs individuals to shape a particular institutional procedure: namely, the bases on which employers select candidates. However, it is not only employers who can violate this requirement (by instituting unfair selection procedures), but also the candidates themselves (by entering unfair selection procedures). Given this, one might think that candidates' families too can contravene this requirement, by providing these individuals with the qualities they need to increase their chances of being unfairly selected. I will admit, I do not see any obvious reason for preferring this way of phrasing the requirement of procedural fairness over an alternative phrasing on which this requirement applies only to employers and legislators. However, I will allow Scanlon the benefit of the doubt and assume the more extensive phrasing.

Even so, I submit that the conflict which Scanlon's view implies between the requirements of procedural fairness and the permissions of parental partiality can be avoided. We should not follow Scanlon, I propose, in thinking that procedural fairness requires parents not to procure for their children benefits which will improve their prospects of being selected through an unfair procedure. This is because there is no absolute standard relative to which attributes count as institution-dependent talents; the same attribute might constitute an institution-dependent talent relative to one type of institution, but not relative to another. As such, if procedural fairness prohibited parents from fostering in their children attributes on the basis of which certain institutions unfairly discriminate, then it could prohibit them from encouraging qualities which constitute institution-dependent talents relative to other institutions. For instance, according to Scanlon, if some institution unfairly requires applicants to hold a master's degree, then procedural fairness requires parents not to use their resources to help their children to acquire one: for instance, by helping to pay their tuition fees. However, it

might be perfectly fair for some other institution to select on the basis of whether or not candidates possess postgraduate qualifications. If so, then Scanlon's position implies that procedural fairness requires parents not to help their children to acquire institution-dependent talents. This, I suggest, is a result we should want to avoid. Provided that the requirements of substantive opportunity are fulfilled, a principle of equal opportunity should not debar parents from encouraging the development of socially beneficial attributes in their children.

An advocate of Scanlon's position might remain unconvinced. 'Of course,' they might respond 'procedural fairness doesn't prohibit parents from helping their children to acquire certain attributes, *so long as* they do so with the intention of helping them to acquire institution-dependent talents. However, it would prohibit them from pursuing those attributes, if they did so with the intention of facilitating their children to fare better in an unfair selection procedures.'

Tweaked as such, Scanlon's view depends on the truth of the Doctrine of Double Effect: the thesis that the permissibility of action can depend on whether its consequences were intended, as opposed to merely foreseen. As such, its plausibility will depend on one's view of that thesis.[36] However, even if one endorses this revised version of Scanlon's claim, it need not lead to conflicts between procedural fairness and a plausible principle of parental partiality. Such conflicts will arise only if the latter principle implies that parents have reasons of partiality to enable their children to fare as well as possible in competitions for advantage, even if those competitions are unfair. Such a partiality principle would be very strong, perhaps stronger than one would be willing to accept.

3.2. School choice

[36] Scanlon's own position is that the Doctrine of Double Effect is false (Scanlon 2008, chap. 1). However, a supporter of Scanlon's views on procedural fairness need not remain faithful to *all* of his philosophical views.

Adam Swift (2003) has argued that, under conditions of social injustice, a principle of equal opportunity would require parents to make decisions about their children's education which could conflict with parental partiality. More specifically, he claims that if the standard of education provided throughout society is unequal, then equality of opportunity requires well-off parents not to use their wealth to acquire a better standard of education for their own children: either by paying for high-quality private tuition, or by moving within catchment of a better-quality state school where property/rental prices are higher.

This, Swift claims, is because education is a *positional good*. A good is positional when the value a person derives from it depends on the access which others have to it. For example, consider the value which a litigant derives from their legal representation. What matters to this person is not how good their lawyer is in absolute terms, but rather how good they are relative to their opponent's lawyer. If their adversary hires better representation, then, regardless of the absolute quality of their own lawyer, they will likely lose their case. As such, the value a litigant derives from legal representation is positional: it depends on the comparative standard of representation to which others have access.[37]

Similarly, part of the value which an individual derives from their education stems from its propensity to improve their prospects of being selected for advantageous positions. Such positions, however, are scarce: there are necessarily less of them than there are individuals who can fill them, otherwise they would not be advantageous. So, the extent to which an individual's education increases their chances of attaining some position of advantage depends not on the absolute quality of their schooling, but rather on its quality relative to that to which others have access. If you and I are competing for the same position, then my prospects of being selected will be

[37] For some helpful discussions of the notion of a positional good, see (Brighouse and Swift 2006; Hollis 1984).

worse if, relative to mine, your education better enables you to develop the qualities on the basis of which candidates for that position are selected.[38]

Given that education is a positional good, when well-off parents use their wealth to acquire a better quality of education for their own child, this will not only procure a benefit for their child. It will also make other children worse off. To see this, consider again the case of the lawsuit. If one litigant hires better representation than their opponent, they do not merely improve their own prospects of success. They also necessarily worsen the prospects of their opponent, since they cannot both win. Similarly, by improving their child's chances of being selected for a position of advantage, wealthy parents do not merely better their own child's prospects. They also necessarily worsen those of certain others: namely, those students whose prospects were previously better than or equal to their child's, but are now worse. As Swift (2003, pp. 24-25) puts it, these students' place in the queue for positions of advantage is pushed back a step.

Given that paying to improve one's own children's education will worsen others' prospects of being selected for advantageous institutional positions, Swift claims that a principle of equal opportunity will prohibit parents from doing so. However, he also notes that a plausible principle of parental partiality will sometimes permit parents to do just that. For instance, if the standard of education which a child currently receives is inadequate, then it's plausible to think that their parents have reasons of partiality to use the resources available to them to provide an adequate standard: even if doing so makes other children worse off (Swift 2003, chap. 8). In such cases, Swift thinks, the requirements of equal opportunity and parental partiality conflict.

However, I suggest that if we understand equality of opportunity in the way set out in the foregoing sections, then it's not clear that, under conditions of educational inequality, this ideal will forbid wealthy parents from purchasing

[38] The positionality of educational goods has also been explored by Martin Hollis (1982).

a better quality of education for their children. As such, the conflict which Swift describes can be avoided.

Which component of Scanlon's three-part conception of equal opportunity would imply such a requirement on parents? Scanlon (2018, pp. 67-69) points out that the positional dimension of education is most acute when procedural fairness is not achieved. If employers did not unfairly discriminate between candidates on the basis of superfluous enrichments, then an education which furnished one with such things would not increase one's chances of being selected for an advantageous position. Parents who moved their children to schools which promise such enrichments would not, therefore, necessarily worsen other student's prospects of advantage; the less enriched children might still retain the same prospects as the more enriched.[39]

Given this, when employers *do* in fact discriminate on arbitrary grounds, thereby amplifying education's positional dimension, one might think that procedural fairness requires parents not to take part in the ensuing educational arms race. More specifically, one might think that it prohibits well-off parents from providing their children with an education which will better enable them to acquire qualities on the basis of which employers unfairly select. However, I argued in the preceding subsection that if employers unfairly discriminate on the basis of certain attributes, then procedural fairness does not forbid parents from helping their children to acquire them. Whilst those attributes might not constitute institution-dependent talents relative to one type of institution, they might relative to another. At most, provided that the Doctrine of Double Effect is true, procedural fairness may prohibit parents from procuring educational benefits for their children *in order* to improve their prospects of being unfairly selected.

[39] Hollis (1982, p. 236) makes a similar point when he claims that one function of the education system is to make hiring choices less costly for employers by enabling them to choose between candidates on the basis of ranked qualifications. Hiring choices will be least costly when employers are able to use qualifications to eliminate as many candidates as possible. As such, if employers hire in this way, then the value of my education as a means for accessing the job market will depend upon the number of other people whose qualifications are similar to, or higher than, mine. However, such hiring practices will also increase the risk of arbitrary discrimination if individuals with lower qualifications would, in principle, be just as competent in a given role as those with higher qualifications.

What about substantive opportunity? Does this component of equal opportunity prohibit well-off parents from paying for their children to receive a higher standard of education than others? Let's grant Swift the claim that, by doing so, a wealthy parent will worsen other children's prospects of attaining an advantageous institutional position: by increasing their child's chances of selection, they will marginally decrease those of others. Even so, it does not follow that substantive opportunity prohibits them from doing this.

To see this, recall what substantive opportunity requires of individuals: it requires us to ensure that the major institutions do not wrongfully block people from attaining institution-dependent talents. Understood in this way, substantive opportunity does not prohibit individuals from worsening others' prospects of being selected for positions of advantage *per se*. Rather, it prohibits a particular way of doing this: namely, by undertaking actions which cause institutions wrongfully to hinder them from acquiring those talents. However, whilst a wealthy parent may worsen others' prospects of being selected for scarce positions by improving their own child's education, they do not thereby worsen others' prospects of developing institution-dependent talents. Those others' chances of selection are worsened not because they have been blocked from attaining the relevant talents, but rather because the pool of competitors (i.e. the pool of individuals with equal or superior talents) has grown. Substantive opportunity permits individuals to worsen one another's prospects in this way. And so it should. If it did not, then it would deliver implausible results. For instance, it would imply that an individual can contravene the requirements of equal opportunity simply by applying for a job: if they were not the only applicant, and had a non-zero chance of being selected, then they would worsen the other applicants' chances merely by applying.

3.3. The family

I noted in Section 2.3 that Scanlon adopts and develops Rawls' stipulation that equally talented individuals ought to have equal prospects of attaining positions of advantage, regardless of their social background. He resolves the

tension within that stipulation, I claimed, by interpreting 'equally talented' in the following way: two individuals are equally talented if and only if, were they to receive an equal standard of education, they would have the same chances of developing institution-dependent talents. However, one might question whether this interpretation is really enough to resolve the tension. More specifically, one might worry that if childcare and initial socialisation take place primarily within the family, then inequalities in people's prospects of acquiring institution-dependent talents will already be fixed before they enter the education system. Different families will value different attributes over others, and so individuals' propensities to pursue and develop institution-dependent talents will differ according to the values and norms embraced by their own family. But if so, then a person's social background will be able to influence their prospects, even under conditions of educational equality.

Motivated by this worry, one might propose the following amendment to Scanlon's definition of 'equally talented': two individuals are equally talented if and only if, were they to receive an equal standard of both education *and initial developmental conditions*, they would have the same chances of developing institution-dependent talents. However, this amendment faces us with a stark conflict between equality of opportunity and parental partiality. Equality of initial developmental conditions would, it seems, require the abolition of the family; so long as some form of this institution exists, standardised childcare and primary socialisation would be impossible to implement. So, if substantive opportunity requires that equally talented individuals have equal prospects of attaining positions of advantage, and 'equally talented' is interpreted according to the amended definition, then it implies that the family ought to be abolished.

Yet parental partiality presupposes that the family, in some form, exists. Principles of parental partiality attach normative significance to the parent-child relationship; they stipulate that if *A* is a parent of *B*, then *A* has reasons to treat *B* in certain preferential ways. There are broadly two ways of construing this relationship. Taking a sociological/anthropological approach,

we could say that the parent-child relationship comprises of certain socially significant forms of behaviour. On this view, we might say that *A* is a parent of *B* when *A* is, or was, a primary caregiver of *B*, or when others treat *A* and *B* in certain distinctive ways. If we assume this sort of conception, then abolishing the family will involve abolishing the parent-child relationship: since the behavioural norms surrounding the family will be constitutive of that relationship. As such, if equality of opportunity requires the abolition of the family, then it will require the eradication of the very relationship to which parental partiality attaches normative significance.

Alternatively, we could assume a biological conception of the parent-child relationship, according to which *A* is a parent of *B* just if the zygote from which *B* developed was formed of a gamete produced by *A*. On this view, the abolition of the family would not result in the abolition of the parent-child relationship; gametes would still form zygotes under some alternative child-rearing system. However, if families were not permitted, it seems doubtful that there would exist any licit opportunities for individuals to express partiality towards their biological children. So long as people are permitted to treat their biological children in preferential ways, one may suspect that the family, in some form, continues to exist. As such, even if we assume a biological conception of the parent-child relationship, the abolition of the family would, in practice, imply the prohibition of parental partiality.[40]

We face a dilemma. The same considerations which led us to think that equality of opportunity requires equality of education also lead us to think that it requires an equal standard of initial developmental conditions for all; only the latter, stronger requirement seems sufficient to ensure that differences in social background will not lead to inequalities in people's prospects. Yet this requirement seems to be incompatible with existence of parental partiality. If we standardise people's initial developmental

[40] It might seem odd to suppose that we have reason to show partiality towards our biological children. Would a parent have reason to show such partiality, even if they had been estranged from their child since birth? For further discussion of this thought, however, see (Kolodny 2010).

conditions, thereby abolishing the family, then we will abolish either the parent-child relationship itself, or the opportunity to express parental partiality.

This potential conflict between equality of opportunity and legitimate parental partiality is, I think, an even greater cause for concern than those described in the preceding subsections. The latter threatened to arise under conditions of unequal opportunity: namely, procedural unfairness in hiring practices, and educational inequality. These injustices need not result from the pursuit of legitimate parental partiality: procedural unfairness could be caused merely by employers' desires to implement cheaper hiring practices; and educational inequality could be the result of disinclination in government, and certain educators, to equalise the distribution of educational goods. Thus, in principle, it might be possible to eliminate those injustices without curtailing parental partiality, thereby dissolving the apparent conflict with equality of opportunity. However, the conflict currently under consideration is different. The point being pressed here is that even if we start from initial conditions of equal opportunity, the pursuit of legitimate parental partiality within the family will inevitably lead to a situation in which people's life prospects are differentially affected by their social backgrounds. As such, there will be no way to dissolve the conflict: we could enjoy equal opportunity only by eradicating parental partiality, and we could protect the expression of parental partiality only by giving up on equality of opportunity. Put starkly, we seem to face a choice between, on the one hand, achieving social justice, and, on the other, relating to our own children other than as strangers.

3.3.1. Rawls' solution

At several points in *A Theory of Justice*, Rawls worries that fair equality of opportunity is incompatible with the existence of the family (Rawls 1999, pp. 64, 447-448). However, he argues that this tension can be mitigated. If the difference principle were to be implemented, Rawls claims, then the need to abolish the family for the sake of equal opportunity would become less urgent. If we assume that people's prospects of attaining positions of advantage will inevitably differ depending on their family backgrounds, then

the difference principle will require that these inequalities work for the benefit of all. That is, assuming that the difference principle has been fulfilled, there will be no alternative way of distributing income and wealth under which the worst off will be better off. As such, those individuals whose family background means that their prospects are worse than others' will be less likely to be downcast about this inequality. Rather, they will look on it as something which works to their advantage (Rawls 1999, p. 448). So, whilst fair equality of opportunity taken in isolation might require the abolition of the family, this becomes much less urgent once we consider the requirements of social justice as a whole.

However, this solution faces a problem. In Rawls' theory, the principle of fair equality of opportunity takes lexical priority over the difference principle (Rawls 1999, p. 266). This means that the former principle must be fully satisfied before the latter can come into operation. Given this, if fair equality of opportunity cannot be achieved whilst the family exists, then the difference principle will never become operant so long as some form of the family prevails. But if this is so, then we cannot appeal to the effects of implementing the difference principle in a justification of the family; so long as equality of opportunity is incompatible with the family, and the family exists, the difference principle cannot be implemented. So, in order for his proposed defence of the family is to work, Rawls must give up the lexical priority of fair equality of opportunity over the difference principle.[41]

3.3.2. Munoz-Dardé's solution

Despite this problem, however, one might maintain that Rawls' general strategy of justifying the family by appeal to the requirements of the principles of justice taken together is salvageable. The problem with his solution is that he tries to rescue the family from the requirements of fair

[41] Véronique Munoz-Dardé (2002, p. 268, fn. 24) argues that the lexical priority of fair equality of opportunity over the difference principle ought to be relinquished because it implies that the difference principle can never become operant so long as the family exists. Whilst this would constitute a fairly substantial revision of Rawls' view, she claims that it would nonetheless retain the spirit of the theory. Arnseon (1999) and Larry Alexander (1986) also criticise Rawls' lexical ordering of the principle of fair equality of opportunity over the difference principle, though for different reasons.

equality of opportunity by appealing to a lexically subsequent principle. But couldn't we simply appeal to a lexically prior principle instead?

This is precisely the approach suggested by Véronique Munoz-Dardé (1999). Munoz-Dardé argues that only the family can provide the kind of moral development required in order for individuals to value equal basic liberty. As such, we can justify the family by appealing to the liberty principle: a principle which, in Rawls' theory, takes priority over equality of opportunity.

If the family were to be abolished for the sake of equal opportunity, Munoz-Dardé asks, what kind of child-rearing institution would replace it? She imagines a kind of generalised, well-run orphanage. Under the orphanage system, children would be separated from their biological parents at birth, and their care would be provided in a standardised way by professionals. These professional carers would not be permitted to give preferential treatment to any child. Rather, they would be obliged to treat all children under their care with impartiality.

Munoz-Dardé argues that, whilst it might bring us closer to achieving equality of opportunity, this alternative child-rearing institution would be unlikely to cultivate individuals for whom liberty is important.[42] In Rawls' theory, the reason why individuals want maximal liberty (and hence why it is rational for parties in the original position to choose the liberty principle) is that it enables them to pursue a diversity of life plans. When childcare and socialisation occur primarily within the family, we can expect such a diversity of life plans to arise, since the lack of standardisation across families will expose children to a range of different norms and values. By contrast, under the generalised orphanage system, socialisation would be standardised; the norms and values

[42] This is not to say that that is the only reason to favour the family over the generalised orphanage system. The development of a close parent-child relationship may well give rise to a number of goods which could not be had in the absence of the family. For an account of such goods, see (Brighouse and Swift 2009; 2014, chaps. 3 and 4). The point, however, is that, given that the family interferes with equality of opportunity, and given Rawls' priority rules for his principles of justice, it will be difficult to make the case that these familial goods are more important than the achievement of equal opportunity unless they can be tied to the liberty principle.

into which children are inculcated would be roughly the same across society. As such, if the family were to be replaced by the orphanage system, we could expect people's life plans to tend, over time, towards homogeneity. This situation, Munoz-Dardé claims, would undermine the rationale which underlies the liberty principle, since it presupposes a diversity of life plans. So, if that diversity disappears, then so too will people's reason for wanting equal basic liberty.[43]

Munoz-Dardé's argument offers a major improvement on Rawls' own attempt to justify the family. Since, in Rawls' theory, the liberty principle is lexically prior to fair equality of opportunity, the requirements of the latter cannot become operant until the former has been fulfilled. So, if the rationale which underlies the liberty principle presupposes some form of the family, then the existence of the family is a pre-condition of equality of opportunity; the family, in some form, must exist in order for the liberty principle to be fulfilled, since its existence is presupposed by the principle's rationale, and so it must exist in order equality of opportunity to become operant. As such, whilst fair equality of opportunity might require the abolition of the family when considered in isolation, it will not require this when considered in combination with the liberty principle.[44]

3.3.3. Shiffrin on priority

Yet, whilst Munoz-Dardé improves on Rawls, there remains a potential difficulty. Munoz-Dardé's solution assumes that Rawls is correct to order the liberty principle as lexically prior to fair equality of opportunity. However, this claim has been challenged by Seana Shiffrin (2004a), who argues that equality of opportunity ought to be ordered on a par with equal basic liberty. If Shiffrin is correct, then it seems Munoz-Dardé's argument will not

[43] I am indebted here to some clarificatory remarks made by Munoz-Dardé in personal communication.

[44] In fact, this argument can be used to supplement Rawls' solution. If Munoz-Dardé is right, then inequalities in people's prospects which are due to differences in their family backgrounds must be admitted for the sake of the liberty principle. However, the difference principle requires that those inequalities must benefit the least well off. As such, if the difference principle is implemented, then individuals may be less disposed to regard the unequal influence of family background on people's prospects with bitterness.

demonstrate that the family is safe from the requirements of equal opportunity. Rather, so long as equal opportunity requires the abolition of the family, it will show only that there is a conflict between the rationale behind the liberty principle, and the requirements of fair equality of opportunity: a conflict which cannot be resolved by appealing to a priority rule.

One possible response here would simply be to defend Rawls' priority rules against Shiffrin's critique. However, I propose an alternative. I submit that if we accept Shiffrin's argument, then a principle of equal opportunity, so conceived, will not require the abolition of the family. As such, we face a happy choice: either Shiffrin is wrong, and so Munoz-Dardé has shown that the family is safe from the requirements of equal opportunity; or Shiffrin is right, and so equality of opportunity does not require the family to be abolished after all.

Shiffrin begins by arguing for a particular rationale for a principle of equal opportunity. Such a principle, she claims, is more than simply a means for enabling individuals to attain positions of advantage, and the benefits attached to those offices. Ensuring that everyone has a fair chance of attaining any position for which they choose to apply enables people to pursue and develop a diverse range of life plans; if I have a fair chance of being selected for any office of my choosing, then I have a fair chance of being able to pursue my own conception of the good, whatever that happens to be (Shiffrin 2004a, pp. 1666-1670).[45] However, Shiffrin points out that this second, more substantive rationale for fair equality of opportunity bears close similarities to that underlying the liberty principle. Viewed in this way, both principles enable people to pursue a diverse range of life plans. As such, if we adopt Shiffrin's rationale for equal opportunity, then it becomes unclear why it should be lexically subsequent to the liberty principle. If both perform very similar functions, then why should one be prior to the other?

[45] In Scanlon's terminology, a rationale which justified equal opportunity as a means for enabling people to attain positions of advantage would be a 'just inequality rationale'. By contrast, Shiffrin's more expansive rationale constitutes a 'self-realisation rationale'. See note 29.

However, if we endorse Shiffrin's claim that the rationales underlying both the liberty principle and a principle of equal opportunity bear close similarities, then there are reasons for thinking that the family, in some form, is a prerequisite of equal opportunity. These reasons will by now be familiar: they are the same are those elaborated by Munoz-Dardé for thinking that the family is presupposed by the liberty principle. The family will be more likely than a general, standardised child-rearing institution to encourage a diverse range of life plans. So, if the rationale behind fair equality of opportunity presupposes diversity of life plans, then a principle of equal opportunity which required the abolition of the family would be self-defeating; it would undermine the social conditions presupposed by its own rationale.

A principle of equal opportunity, of the sort envisaged by Shiffrin's expanded rationale, would thus stop short of requiring equality of initial developmental conditions. That is, such a principle would permit some inequalities in people's prospects of attaining positions of advantage to arise as a result of differences in family background. However, this is not a concession to some further, lexically prior principle. Such inequalities would rather be consistent with the requirements of equality of opportunity. This may seem strange, given the idea discussed above that substantive opportunity requires that people's prospects not differ according to social background. However, it will seem less strange if we keep two points in mind. First, given Shiffrin's rationale, a principle of equal opportunity will still require equality of education. An equal standard of education for all will dampened the differential effects which family background has on people's prospects. Second, if the point of equality of opportunity is less to enable individuals to compete fairly with one another for wealth, and more to enable them to pursue a diversity of life plans, then permitting small inequalities to arise as a result of differences in family background will not seem so contrary to this ideal. If such inequalities are an inevitable result of maintaining a diversity of norms and values, then equal opportunity, so conceived, will require us to choose diversity.

3

Structural Injustice

The guiding question of Chapters 1 and 2 was this: What do Rawls' principles of justice require of individuals? In this Chapter, I shift my focus from Rawls' theory to consider a concept which has recently received much attention in political theory: the notion of structural injustice. The focal point of this attention is the work of Iris Marion Young; whilst she did not coin the term, Young has done much to develop this concept, and to consider some of the philosophical questions which it raises. Taking Young's work as my starting point, my aim in this chapter is to reach a clearer understanding of the idea of structural injustice, and of what we gain by introducing this concept into a theory of social justice; Sections 1 and 2 take up these two aims respectively. I also include an appendix, in which I argue that, at least on one influential account of harm, social injustice can be considered a harm. As well as being implicitly assumed at several points in the chapter, the idea that social injustice is a harm will help to bridge the gap between the subject of Chapters 1, 2, and 3, and that of Chapters 4 and 5.

1. What is structural injustice?

Addressing scepticism about the study of social structure in anthropology, Claude Lévi-Strauss quotes A. L. Kroeber:

> 'Structure' appears to be just a yielding to a word that has a perfectly good meaning but suddenly becomes fashionably attractive for a decade or so … and during its vogue tends to be applied indiscriminately because of the pleasurable connotations of its sound … So what 'structure' adds to the meaning of our phrase seems to be nothing, except to provoke a degree of pleasant puzzlement (Kroeber 1948, p. 325, quoted in Lévi-Strauss 1963, p. 278)

A similar concern might be directed towards theorists who employ the notion of structural injustice. What, exactly, does the adjective 'structural' add to this phrase? Does it designate a subset of the set of all injustices? If so, what distinguishes the 'structural' instances of injustice from the 'non-structural' ones? And if not, is it nonetheless informative to describe injustice as 'structural', or does the phrase become redundant? My aim in this section is to go some way towards answering these questions, taking as my focus the conception of structural injustice employed by Young.

First, however, it will be instructive to consider some examples of the phenomenon which Young intends to characterise. Her work is illustrated with a broad range of examples which include: the persistence of sweatshop conditions sustained by consumer demand for cheap apparel (Young 2004; 2006a; 2011, chap. 5); the effect of the organisation of domestic labour on economic inequality between men and women (Young 2001, pp. 10-11); environmental damage caused, in part, by increased car ownership (Young 2011, p. 103); the disadvantage faced by Black and Latino communities in the aftermath of hurricanes Katrina and Rita (Young 2006b); and the ongoing legacy of American slavery (Young 2011, chap. 7).[46]

Perhaps the most widely discussed of Young's examples, however, is that of Sandy (Young 2011, pp. 43-44). Sandy is a single mother who is given notice to vacate her apartment when her landlord decides to sell the building to a developer. When she begins to search for a new apartment, she discovers that most of the properties close to her workplace are outside of her price range. Settling for an apartment further away, she uses her savings for a down payment on a car. However, with her eviction deadline imminent, she then learns that she must pay three months' rent upfront in order to secure a deposit on a flat: money which she has already spent on the down payment. With

[46] Young's discussion of historic injustice has been developed further by Catherine Lu (2011) and Magali Bessone (2019).

nowhere else to turn, Sandy and her two children face the prospect of living out of her car.

The case of Sandy seems evocative. But evocative of what? More precisely, what feature does this case share with the others mentioned above which marks it out as an instance of *structural* injustice?

1.1. Structural injustice as social injustice

One possible answer runs as follows. Young claims that it is intuitive to think that Sandy, in facing the prospect of homelessness, suffers a wrong. However, the wrong which she suffers, Young thinks, is of a distinctive kind. More specifically, it is distinct both from the kind of wrong which can occur in individual interaction (cases in which one person 'wrongs' another), and from that which can occur in an individual's interaction with the state (cases in which a specific state policy sanctions the wrongful treatment of one person by another). According to Young, Sandy has not been treated wrongfully by another individual. Nor is her situation the result of any one state policy which authorises wrongful treatment. As such, the wrong she suffers must be of some third kind: a kind which we might refer to as 'structural' (Young 2011, pp. 45-47).[47]

This is a step in the right direction. But it does not yet give us a positive characterisation of structural injustice. We merely have a statement of what it is *not*: instances of structural injustice are equivalent neither to instances of wrongful personal interaction, nor to instances of wrongful state treatment. What we are lacking is a positive account of the distinctive kind of wrong which individuals like Sandy are supposed to suffer. In other words, Sandy, we are assuming, has a legitimate complaint. But her complaint is not that she has been treated wrongfully by someone else. What, then, is her complaint about?

[47] Lu (2018, pp. 45-47) emphasises the distinction between 'interactional' and 'structural' wrongs in her interpretation of Young's notion of structural injustice.

The discussion of Chapter 1 brings one possibility to mind. There, we saw that Rawls distinguishes between justice understood as a virtue of individuals and justice construed as a virtue of social systems. Social justice, for Rawls, is a matter of the distribution of fundamental right, duties, and benefits determined collectively by society's major institutions; a society is more just to the extent that this distribution meets the specifications of the principles of justice, and is less just to the extent that it does not. When a society in unjust, those who suffer this injustice[48] can legitimately complain of their situation. But this complaint is not one which concerns something which another person has done to them. Rawls makes this point clearly when he writes:

> The fact that everyone with reason believes that they are acting fairly and scrupulously honoring the norms governing agreements is not sufficient to preserve background justice. This is an important though obvious point: when our social world is pervaded by duplicity and deceit we are tempted to think that law and government are necessary only because of the propensity of individuals to act unfairly. But, to the contrary, the tendency is rather for background justice to be eroded even when individuals act fairly (Rawls 1993, p. 267)[49]

An individual can legitimately complain that they suffer social injustice, even if they cannot complain about the way in which they have been treated by others. Rather, for Rawls, their complaint is that the basic structure of their society is arranged in a way that denies them the protections, opportunities, or benefits which, as a member of society, they are owed. As such, the victims of injustice have a claim on others to improve their situations by reshaping the basic structure, even if none of those individuals have treated them wrongfully.

[48] More specifically, those whose basic liberties are denied or not protected, those who are denied a fair chance of attaining the various positions on offer, or those disadvantaged by inequalities which do not benefit the worst off.
[49] Young (2011, p. 73) herself approvingly misquotes this passage.

Given this, we might identify the kind of wrong suffered by Sandy, and others like her, with the kind of wrong described by Rawls. Whilst Sandy cannot complain of having been treated wrongfully by other people, she can complain that she has not received her due: perhaps a fair chance of being selected for positions of advantage, or else the benefits which she would receive if all socio-economic inequalities were to the greatest benefit of the least well off. Understood in this way, the distinctive kind of wrong which Young aims to isolate will be equivalent to Rawlsian social injustice.[50]

What should we make of this way understanding the notion of structural injustice? Viewed as such, the concept draws attention to an important idea: that individuals can have claims on others to improve their situation, even if they have not been treated wrongfully. Yet the interpretation is nonetheless disappointing. Presumably, advocates of the concept would hope that, by describing something as an instance of structural injustice, they are saying something more than that it is merely unjust in Rawls' sense. That is, they might hope that the notion designates some subset of the set of all social injustices. Alternatively, if the terms 'structural injustice' and 'social injustice' do turn out to be coextensive, they might nonetheless hope that their meanings differ in some other way, such that it remains informative to describe an instance of social injustice as 'structural'.

One might, however, resist this verdict. Granted, Rawlsian social injustice can occur without any identifiable perpetrator; and in such cases we can say that the victims of injustice have a legitimate complaint which does not concern the way in which they have been treated by others. But an individual could also be placed in a situation of apparent injustice by the wrongful actions of another. For instance, I might be substantially worse off than others

[50] Tommie Shelby (2016, p. 28) distinguishes one sense of the notion of structural racism which accords with Rawls' concept of social injustice. Structural racism, in this sense, occurs when the policies of several institutions, which may not themselves be designed to disadvantage any racial group, collectively have disproportionately adverse effects on members of a particular racial group. Shelby notes that the kind of complaint which individuals have against structural racism, so construed, is the same as that which individuals have against social injustice more generally: namely, that the basic structure does not provide them with what they are due.

because someone else has stolen the holdings to which I am entitled under a justified scheme of private property. If I am thereby made worse off than I would be under an equal distribution, then it seems that the difference principle will imply that I can legitimately complain of my situation. Given this, there may be space to differentiate structural injustice from Rawlsian social injustice more generally: injustice will be structural whenever it has no identifiable perpetrator(s).

This proposal goes beyond that so far explored in this subsection by introducing issues of aetiology. The idea I have been considering is that structural injustice is a distinctive kind of wrong, one which differs from the wrongs of personal interaction. My argument has been that Rawlsian social injustice constitutes a wrong of precisely this sort. As such, if structural injustice is construed in the way under consideration, then it can be assimilated with Rawlsian injustice. The above response concedes that structural injustice might be a *type* of Rawlsian injustice, but maintains that it can be distinguished from other types of injustice by virtue of the way in which it has been brought about. We have thus moved on from characterising structural injustice as a distinctive kind of *wrong*, to thinking of it as a category of injustice which can be *explained* in a distinctive kind of way. I will explore this proposal in more detail in the following subsection. However, before I do so, I want to express some scepticism about the response given in the preceding paragraph.

More specifically, I want to voice doubt about the suggestion that I have a claim of social injustice in the event that I am made worse off than others as the result of theft. This, I have said, would amount to a claim on others to reshape the basic structure of society so that it provides me with the benefits I am due. However, this seems to misplace the moral claim that I have in such situations. It seems more natural to think I have claim on the *thief* to return my property, or at least to compensate me for my loss, than it is to think that I have claim on my fellow citizens to reform the basic structure so that it makes me better off.

We can theorise this judgement as follows. When an individual incurs a loss as the result of another's wrongful behaviour, they have a claim on that person to recover that loss.[51] Moreover, when an individual possesses a valid claim of this sort, it pre-empts any claim of social justice which they might otherwise have on their fellow citizens to improve their situation up to the level which would be achieved by the recovery of their losses. As such, in cases where a person has been made worse off by someone else's wrongful action, if recovering their losses would mean that they could not legitimately complain about their situation according to the principles of justice, then they have no complaint of social justice; that social-justice complaint is overridden by a distinct claim to restitution.

Notice that this suggestion can be neatly accommodated within Rawls' institutional division of labour (see Chapter 1, Section 2.3). The precepts of private law are designed to adjudicate the sorts of moral claim which arise when an individual incurs a loss due to the wrongful behaviour of another. By contrast, the principles of social justice (or at least, those which concern the distribution of wealth) are to be enforced through a different sort of regulation: one which allocates resources not from a particular wrongdoer to the person they have wronged, but rather from the citizen body in general to those who do not receive their fair share of the benefits of social cooperation. What I am suggesting is that in cases where a claim of the sort dealt with by the private law coincides with those dealt with by the latter sort of regulation, the former takes precedence.

If this is right, then the conception of structural injustice proposed a few paragraphs ago will face a difficulty. According to that conception, non-structural instances of injustice are those which occur as the result of a discrete piece of wrongful behaviour. However, if the thought presented in the preceding two paragraphs is correct, then a situation in which someone suffers a loss because of someone else's wrongful conduct will not be an instance of social injustice; not unless they would retain a social-justice

[51] See note 61.

complaint even after restitution. And even in those cases of wrongful loss where the victim *does* have an un-pre-empted social-justice claim, it would be misleading to describe them as cases in which injustice is caused by a given individual's wrongful action; the wrongdoer makes the victim's situation worse, but it was already unjust before the loss was incurred. As such, even if they do not disappear entirely, non-structural instances of injustice will be a rarity, and so the hoped-for distinction between structural injustice and social injustice more generally will become difficult to maintain.

1.2. Structural explanation

However, as I have already gestured, there is another way of understanding Young's concept. Rather than characterising it as a distinctive kind of wrong, this alternative focusses instead on the distinctive aetiology of structural injustice. We have seen already that, according to Young, instances of structural injustice arise neither as the result of how one person treats another, nor as the consequence of actions sanctioned by a particular state policy. How, then, do these injustices come about? In order to explain instances of structural injustice, Young claims that we need to appeal to what she calls "social-structural processes" (Young 2011, p. 53). Here, Young is making use of the social-scientific concept of a *social structure*; the basic thought is that, since they cannot be explained by reference to a single individual or policy, instances of structural injustice are to be explained in terms of social structures.

This claim can be used to offer a positive characterisation of the notion of structural injustice itself. According to this characterisation, structural instances of injustice can be distinguished from non-structural instances by reference to their aetiology: structural injustice is explained structurally (that is, by appeal to social structures), whilst non-structural injustice is explained in some other way.[52]

[52] Andrea Sangiovanni (2018, p. 463) endorses this way of understanding Young's concept of structural injustice. He distinguishes it from some alternative ways of construing the idea which tie it more closely with Rawls' notion of social injustice.

This way of understanding the concept appears to have a number of advantages over that presented in the preceding subsection. First, on this characterisation, the notion of structural injustice will not necessarily be coextensive with that of Rawlsian social injustice, unless it turns out that social injustice can never adequately be explained without reference to social structures (more on this shortly). But second, even if these two concepts did turn out to be coextensive, it would not be uninformative to describe an instance of social injustice as structural. To do so would be to say something about the aetiology of that social condition, whereas to say baldly that it is unjust would be to say that the individuals in question have a distinctive kind of claim on others to improve their situation. And finally, as a bonus, the account makes good sense of Young's choice of the adjective 'structural'; applying this adjective to the noun 'injustice' is not an indiscriminate yielding to a passing fashion, but is rather an attempt to say something about how unjust social conditions arise.

Yet interpreting the notion of structural injustice in terms of aetiology also brings difficulties. In order to attach any substantive meaning to it, we will need explications of the concept of social structure, and of the distinctive role it plays in the explanation of social phenomena. However, this concept is one of the most deeply contested in the social sciences.[53]

Of course, Young was aware of this. She avoids committing herself to any explicit definition of social structure, instead restricting herself to outlining four general features of the concept's denotation which social scientists typically emphasise: (1) social structures are experienced as objective features of the world which constrain and enable people's options; (2) social structures can be conceptualised as an abstract space of interrelated 'positions'; (3) social structures are produced and reproduced by individual action; and (4) the outcomes produced by social-structural processes are typically unintended (Young 2011, pp. 53-64).

[53] For a helpful discussion of the concept of social structure, and of some of the debates surrounding it, see (Sewell 1992)

These four features are illuminating. But they do not provide a complete answer to our question: How is the kind of explanation which appeals to social structures differentiated from other forms of social-scientific explanation? Without an answer to this question, construing structural injustice in terms of structural explanation will not in fact yield the hoped-for advantages mentioned above. That is, it will not yield a clear differentiation between the structural and non-structural forms of injustice, nor will it be clear what we learn by discovering that an instance of injustice is structural.

Happily, however, there is a popular account of social-structural explanation which seems to cohere with Young's discussion. On this view, a social structure is an abstract system of 'positions', 'offices', or 'nodes'. These abstract positions are defined by the relations that hold between individuals in the concrete social system which the structure maps. A passage in Italo Calvino's *Invisible Cities* provides a haunting illustration of this idea:

> In Ersilia, to establish the relationships that sustain the city's life, the inhabitants stretch strings from the corners of the houses, white or black or gray or black-and-white according to whether they mark a relationship of blood, of trade, authority, agency. When the strings become so numerous that you can no longer pass among them, the inhabitants leave: the houses are dismantled; only the strings and their supports remain.
>
> …
>
> Thus, when traveling in the territory of Ersilia, you come upon the ruins of the abandoned cities, without the walls which do not last, without the bones of the dead which the wind rolls away: spiderwebs of intricate relationships seeking a form (Calvino 1997, p. 68)

A social structure, on the present view, is like an abandoned Ersilian city, a spiderweb of relations without any of the people whom they relate. The positions within this structure represent the gaps where the people used to be: or, more precisely, the points in the abstract structure which are occupied by individuals in the corresponding concrete social system.

Individuals in the concrete system face limitations on the options they can pursue by virtue of the relations in which they stand to others. As such, we can map these limitations onto the abstract positions within the social structure. For example, if a person occupies the position of 'tenant' in a land-tenure system, then, in virtue of occupying that position, their opportunities will be restricted in certain ways: they will not be able to gain access to land without paying rent to a landowner, for instance. Given this, we can explain the behaviours which people engage in, and the events which result from such behaviour (revolutions, wars, technological innovations), in terms of the options which their structural positions make available.[54]

Assuming this construal of structural explanation, the notion of structural injustice will designate instances of injustice which can be explained by appealing to the range of options attached to different structural positions. Sandy, for instance, might face a limited range of options by virtue of occupying certain structural positions: woman, mother, tenant etc. And the others around her might find their own options enabled or restricted by the positions which they occupy. Delineating these option-sets might help us to explain why individuals like Sandy face the prospect of homelessness.

Notice, however, that on this view, structural explanation does not constitute a *sui generis* type of explanation. Rather, it constitutes an elaboration of a more fundamental form of explanation: namely, causal explanation, and in particular the variety of causal explanation expounded by rational choice theory. Put roughly, causal explanations function by positing a causal

[54] For views of this kind see (Hanslanger 2016; Little 1991, chap. 5; Satz and Ferejohn 1994). Daniel Little's chapter offers several illuminating examples of this form of explanation drawn from social-scientific research.

mechanism which leads from the occurrence of the *explanans* to that of the *explanandum*. Rational choice theory posits a particular kind of causal mechanism between *explanans* and *explanandum*: it offers a paradigm of rational decision making.[55] Given this paradigm, the theory says that if we can assume that individuals will choose rationally, and we can give sufficiently detailed descriptions of their environments, then we can explain various social phenomena as the aggregate results of the rational choices of large numbers of individuals. On the view I am currently assuming, structural explanation simply offers a description of individuals' environments which can help us to determine the rational choices open to them. As such, appeals to structural positions are not, on their own, sufficient to explain social phenomena. Rather, if they are relevant to the explanation of some phenomenon, it is only as a component of a larger causal model.[56]

This introduces a complication. One of the hoped-for advantages of construing the notion of structural injustice in terms of structural explanation was that doing so would provide a clear distinction between the structural and the non-structural instances of injustice: the structural injustices are those which are explained structurally, whilst the non-structural ones are those which are not explained in this way. However, if structural explanation is not a *sui generis* form of social-scientific explanation but is rather a part of a wider causal model, then this distinction will be difficult to draw. To say that a social condition is explained structurally will not be to say that it is explained in a unique kind of way, distinct from other forms of explanation. Rather, it will be to say that it is explained causally, where the relevant causal model takes into account information about the restricted option-sets attached to various social positions.

[55] For a classic defence of the view that rational explanations of action are a species of causal explanation, see (Davidson 1963).
[56] See (Little 1991, pp. 103-106). Debra Satz and John Ferejohn (1994, pp. 83-84) argue that it can be appropriate for an explanation of some social phenomenon to end at the structural level: that is, in a delineation of the options and interests typical of certain structural positions. The appropriate endpoint of an explanation, they claim, will depend on the question which is being asked by the theorist. For a critical response, see (Hausman 1995).

But if this is so, then what will it mean to say that a social condition is *not* explained structurally? Either that it is not explained by a causal model of the sort employed in rational choice theory, or that the rational-choice-theoretic model which explains it does not employ information about social positions. However, one might be sceptical about whether either of these disjuncts yields an adequate form of social-scientific explanation. On the one hand, one might think that rational-choice explanations are foundational in the social sciences, and that explanations of social phenomena which cannot be construed in terms of rational choice are ultimately untenable.[57] On the other, one might doubt whether causal models which do not employ information about the option-sets attached to structural positions are ever sufficient to explain complex social phenomena.[58]

The upshot of all this is that it may very well turn out that there are no, or at least very few, instances of social injustice which are not explained structurally. But if so, then the concepts of structural injustice and social injustice will be more or less coextensive. As such, the first hoped-for advantage of construing structural injustice in terms of aetiology might turn out to be chimerical. However, the second hoped-for advantage may still be tenable. Even if it turns out that most, if not all, social injustices are explained structurally, it may nonetheless be informative to draw attention to this fact. Thus, even if the notion of structural injustice is approximately coextensive with that of social injustice, it may not be redundant.

2. The significance of structural injustice

The preceding section explored two ways in which we might try to define Young's notion of structural injustice. According to the first, it is a distinctive kind of wrong which cannot be explicated in terms of wrongful personal interaction. According to the second, structural injustice is social injustice

[57] See e.g. (Elster 1982).

[58] Satz and Ferejohn (1994, pp. 78-81) argue that rational choice theory is predictively accurate only under conditions in which choice is heavily constrained. This, they claim, shows that rational-choice explanations in the social sciences are most credible when supplemented with descriptions of social-structural constraints.

which is explained structurally. My aim in this section is to establish the philosophical import of Young's concept.

If we adopt the first construal of structural injustice, then it seems the concept's import will be nil; it is simply a new name for an old idea: namely, Rawlsian social injustice. However, the second way of characterising the concept seems more promising. Understood in this way, by describing an instance of injustice as structural, we are indicating something about its aetiology: namely, that it can be partially explained in terms of the option-sets which attach to various social positions. In what follows, I will assume this second construal of the concept. My question, therefore, will be this: What, if anything, is contributed to a philosophical theory of justice by recognising that injustice can be explained structurally?

2.1. Revealing injustice

According to some theorists, structural explanation can reveal injustice. If we attempt to explain certain unjust social situations without appealing to structural constraints, these theorists claim, then we will arrive at the erroneous conclusion that those situations are just. In order to avoid such errors, we need to employ structural explanation. Thus, on this view, the contribution which the concept of structural injustice makes to the theory of justice is epistemic. What we gain by adopting it is a more accurate picture of the extent to which injustice permeates our societies; without it, our theory will obfuscate injustice.

Young herself has advocated this view. She claims that when forming judgements about the justice or injustice of social conditions, it is crucial that we attend to how those conditions are explained, and in particular to whether they can be explained in a structural way (Young 2001, pp. 6-9). However, in this subsection, I will concentrate on an argument put forward by Sally Haslanger (2016). Haslanger retains the ideas which seem to motivate Young. Yet, unlike Young, she organises them into an explicit argument for the conclusion that structural explanation reveals injustice.

Haslanger illustrates her argument by focussing on the case of economic inequality between men and women. Such inequality, she claims, is typically unjust; however, we can only see this when it is explained structurally: that is, when we explicate women's relative disadvantage by pointing out that they are positioned within a social structure which constrains their options in certain ways. By contrast, if we explain this inequality in an individualistic way, then we will be forced to conclude that it is just (Haslanger 2016, pp. 121-125).

Before addressing Haslanger's argument in any further detail, I want to note that it relies on an uncertain distinction between structural and individualistic explanation. According to the account presented in Section 1.2, structural explanation is not a *sui generis* form of social-scientific explanation but is rather an incomplete element of a wider causal model. Individualistic explanation, I suggest, can be understood in much the same way. As Haslanger (2016, pp. 121-122) construes it, to explain some social phenomenon individualistically is to demonstrate that it is caused by the mental states of certain individuals. This is precisely the sort of claim which is supposed to supplement structural explanations in the wider causal model. However, notice that the supplementation can go the other way as well. That is, whilst a structural explanation will be insufficient on its own to explain some social phenomenon, it may well be that a solo individualistic explanation is unlikely to fare much better; unaccompanied by information about the constraints on choice attached to certain structural positions, a given set of mental states might be equally likely to lead to any number of possible outcomes.[59] But if this is correct, then it will be a mistake to think of structural and individualistic explanations as distinct forms of explanation which can be employed independently of one another. Rather, they will both be mutually reinforcing parts of the same causal model. Applying the point to the case at hand, considered on their own, neither a structural nor an individualistic explanation of gendered economic inequality will be adequate; only when

[59] See note 58.

combined within a larger causal model will they suffice to explain this phenomenon.

For the sake of argument, however, let's suppose that individualistic and structural explanation can be separated out in a satisfactory way. Why think that one of these forms of explanation obscures injustice, whilst the other reveals it? Haslanger writes:

> Given only the … individualistic [explanation], the fact that women remain economically disadvantaged relative to men appears not to be a matter of moral or political concern: if the best explanation of women's choices to forego economic success is that they, as individuals, desire to be caregivers of children (and the elderly), this is a choice we must respect. No intervention in the name of justice is called for (Haslanger 2016, p. 124)

Haslanger's reasoning seems to run as follows. An individualistic explanation of women's economic disadvantage will posit that, typically, women prefer to care for family members over pursuing a career, and so choose to forgo the economic benefits of the latter in order to pursue the former. However, if someone is worse off than others as a result of their own choices, then their situation is not unjust: that is, they cannot legitimately complain of being worse off than other people. So, if gendered economic inequality is explained individualistically, then it must be just; since women's relative disadvantage will be the result of their own choices, they will have no legitimate claim on others to improve their situation.

This argument assumes what I will call *choice-sensitivity*: the claim that if a person is worse off than others because the choices they have made, then they cannot legitimately complain of their relative disadvantage. Choice-sensitivity has been the subject of much debate over the last few decades. It is endorsed by the family of views gathered under the label 'luck

egalitarianism'; and this aspect of the luck egalitarian view has, in turn, been roundly criticised.[60]

On which, if any, side of this debate does the theory of social justice which I have assumed and developed in the preceding chapters fall? Notice that I have already discussed something close to choice-sensitivity. In Chapter 2, I considered Rawls' claim that equally talented and motivated individuals ought to have equal prospects of attaining positions of advantage, regardless of their initial place in the social system. Following Scanlon, I distinguished two ways of interpreting this claim. On the first, Rawls' claim implies that people's prospects should *always* be permitted to differ according to how they choose to employ their talents. On the second, by contrast, it implies only that people's prospects should *sometimes* be allowed to differ according to the choices which they make. The first reading equates to a bounded version choice-sensitivity. It says that, given that *A* and *B* are equally talented, *A* cannot legitimately complain of being worse off than *B* if *A*'s relative disadvantage is the result of choices which she has made. The second reading, however, gives an even more restricted version of choice-sensitivity. On this reading, even if *A* and *B* are equally talented, and *A* is worse off than *B* because of her own choices, *A* might still have a legitimate complaint. This will be so if *A*'s situation is such that she does not have reason to want what happens to her to depend on how she chooses.

Recall that, following Scanlon's lead, I suggested that we reject the first reading of Rawls' stipulation, since it leads to a tension between its two parts. Instead, we should adopt the second reading, and say that people's prospects should be allowed to differ according to their choices only insofar as their choices have value. As such, the conception of equal opportunity advocated in Chapter 2 implies a very restricted version of choice-sensitivity. And, moreover, this conception is only a part of a wider theory of justice which places further restrictions on choice-sensitivity. In Rawls' theory, the

[60] For some classic expositions of choice-sensitivity, see e.g. (Arneson 1989; Cohen 1989; Dworkin 1981). For criticism, see e.g. (Anderson 1999; Scheffler 2003; Shiffrin 2004b).

requirements of equal opportunity are supplemented by the difference principle. As such, even if equality of opportunity permits *A* to be worse off than *B* because of the choices which she has made, *A* may nonetheless have a legitimate complaint grounded in the difference principle.

The theory of justice which I have adopted here thus rejects the strong version of choice-sensitivity assumed by Haslanger. Furthermore, if we substitute that strong claim for the very weak version of the claim implied by this theory, then her argument does not go through. Even if we explain gendered economic inequality solely in terms of the choices which women make, we do not get the result that this inequality is just. Assuming that everyone has already been secured equal basic liberty, we need to establish two further things. First, we must establish whether the relevant choices were made in conditions under which they have value. Second, we will need to establish whether the inequality in question is most beneficial to the least well off. Without such further information, the fact that women's relative disadvantage is the result of their own choices is insufficient to tell us that "no intervention in the name of justice is called for".

So, if we assume the account of justice adopted in the preceding chapters, an individualistic explanation of gendered economic inequality will not imply that such inequality is just. But what about structural explanation? Is this form of explanation sufficient to ground a judgement of injustice? Haslanger writes:

> The structural explanation reveals, however, that there is a deeper problem than the wage inequity. The … explanation shows that women as a group are structurally situated so that it is rational for them to choose options that keep them subordinate … Without the structural explanation, injustice is obscured (Haslanger 2016, p. 124)

Haslanger highlights something which it is important for theorists of justice to recognise: in virtue of occupying certain structural positions, many women

face choice-situations which make it rational for them to make choices which result in their being disadvantaged relative to men. I will return to this point later. However, I suggest that its significance is not that it is sufficient to warrant a judgement of injustice. On Rawls' theory, an inequality could be just even if it is caused by choices made rational in part by the structural constraints people face. This will be so if those constraints satisfy the requirements of equal opportunity, and the inequality is to the advantage of the least well off. Perhaps there are assumptions that Haslanger could introduce which would validate her inference. But to do so would be to beg the question against theories of justice which do not make such assumptions.

2.2. The requirements of justice: social connection

So much for the view that structural explanation reveals injustice. There is, however, an alternative way of construing the significance of structural injustice. On this view, acknowledging that injustice can be explained structurally does not give us a more accurate picture of the extent of social injustice, but rather extends the responsibilities which individuals bear with respect to injustice. Once we introduce the concept of structural injustice into our theory, advocates of this view allege, we will be forced to recognise that justice requires much more of us than we had previously thought.

This is the approach which Young develops most explicitly. She argues that we need a theory which will tell us what our responsibilities are in relation to structural injustice. The dominant model of individual responsibility assumed by legal and ethical theory, she claims, is inadequate for this task. Young refers to this as the *liability model*. According to this model, a person is assigned responsibility for some harm if and only if their behaviour caused the harm in question, and they were at fault for behaving as such (Young 2011, pp. 97-98). Responsibility here is understood in what Scanlon (1998, p. 248) calls its "substantive" sense: that is, as moral requirement. So, to assign a person responsibility on the liability model amounts to the claim that they

owe something to someone else because they wrongfully caused harm to them.[61]

Young argues that the liability model is unable to assign people responsibility for the harms of structural injustice. This is because such harms cannot be traced to the behaviour of any one individual, or even a small group of individuals. Structural injustice results from largescale behavioural patterns, and the harmfulness of any particular action within that pattern is typically impossible to determine (Young 2011, pp. 99-100). This claim, that structural injustice arises and persists as the cumulative result of many individuals' actions, coheres with the understanding of the concept which I have been assuming. Indeed, it is implied by the account of structural explanation given in Section 1.2. To explain some injustice *I* structurally, on this view, is to delineate the options which attach to certain structural positions. To do this is, in turn, to give a description of individuals' environments which can help to identify the rational choices open to them at a given time; and through identifying these choices, we can offer a causal explanation of *I* by describing it as the aggregate result of the rational choices of a large number of individuals.

In response to the liability model's inability to assign individuals responsibility for structural injustice, Young develops a new model of responsibility which she labels the *social connection model*. Unlike the liability model, Young's new model does not assign an individual responsibility for a harm just if their culpable action can be isolated as *the* cause of the harm. Rather, on the social connection model, a person is assigned responsibility for a harm if their action is merely a *part* of what causes it: that is, if it is one among many things which, taken together, cause the harm.[62] Furthermore, an individual need not be at fault in order to be

[61] This claim raises questions which are passed over by Young: *Why* do I owe you something if I have wrongfully caused you harm? What grounds this duty? And, assuming that I do have this duty, *what* do I owe you? For a recent discussion of these, and other related, questions, see (Gardner 2018b).

[62] Young relies on an intuitive understanding of partial causation. However, giving a rigorous account of this notion presents a significant theoretical challenge. See (Goldman 1999) for discussion.

assigned responsibility on the social connection model; if an individual non-culpably contributes to a causal process which brings about a harm, then this is sufficient for them bear substantive responsibility with regard to it. Finally, the moral requirements which the social connection model assigns are of a particular type. Individuals are not, for instance, required to pay compensation to those harmed. Rather, if they are assigned responsibility on the social connection model, then they are required to join together with others to transform the causal processes which lead to harm (Young 2011, pp. 104-113).

It should now be clear that where the liability model does not assign responsibility for structural injustice, the social connection model will. According this model, all individuals whose actions causally contribute to some instance of structural injustice bear responsibility for changing that causal process. As such, if acknowledging the existence of structural injustice forces us to supplement the liability model with the social connection model, then a theory which recognises that injustice can be structural will posit more extensive individual responsibility for justice than a theory which adopts only the liability model.

I have two points to raise in criticism of Young's view: the first is a direct criticism of her social connection model of responsibility, and the second concerns the implications of her argument for Rawls' theory. To begin with the direct criticism, Young's model assumes that partial causation of a harm is sufficient to ground substantive responsibility to repair that harm. This claim, I submit, delivers unintuitive results. For example, suppose that a hitman is pursuing someone who looks to him almost exactly like his target. In fact, he is mistaken; the individual he is pursuing is entirely unconnected with the hit. However, as luck would have it, this individual leads him to the location of the actual target. As the fake target disappears from view behind a hedgerow, the real target emerges; and as they emerge, they are shot and injured by the hitman. The fake target's actions, it seems, are a partial cause of the real target's injury; the former's behaviour caused the hitman to be in a position to shoot the latter. But do they thereby bear responsibility to

mitigate the harm caused? It seems odd to think so. If anything is owed to the injured victim, it is owed by those who planned, ordered, and carried out the assassination attempt.

Supporters of Young's view might respond by claiming that this is the sort of situation in which the liability model, rather than the social connection model, ought to be employed to mete out responsibility. Young is clear that she does not want to do away with the liability model. Rather, she claims that it requires supplementing in cases where harm cannot be traced to a discrete set of actions. Since this is not such a case, there is no need to defer to the social connection model; the liability model is adequate to identify the individuals who bear responsibility for mitigating the harm in question.

However, this response misses the point. Young's social connection model tells us that if an individual non-culpably performs an action which is a partial cause of a harm, then they bear responsibility for mitigating that harm. The point of the hitman example is to demonstrate the implausibility of that claim; considered in isolation, the idea that non-culpable, partial causation of a harm can ground a responsibility to mitigate it seems unintuitive. But if this is correct, then the social connection model does not get off the ground. As such, the claim that the prescriptions of this model are superseded by those of the liability model in certain cases is beside the point; if non-culpable, partial causation cannot ground responsibility to mitigate harm, then the social connection model cannot plausibly be invoked in *any* case. To put the point another way, if there are cases in which it is intuitive to think both that an individual's action is a partial cause of a harm, and that they bear responsibility to mitigate it, then their responsibility must be grounded not in their causal contribution but rather in some further feature of the case.[63]

For this reason, I suggest that Young's social connection model is untenable. Notice, however, that this need not commit us to the liability model. Indeed, as a general theory of when an individual bears responsibility for mitigating

[63] See (Sangiovanni 2018, p. 468) for a similar criticism of Young's social connection model.

a harm, this model also seems inadequate. There are a number of reasons why a person might bear such responsibility, even if it is not a result of culpable behaviour on their part. For example, they might simply be in a position to provide help; or they might occupy a role which obliges them to assist (a position in the emergency services, perhaps); or they might have promised to pay compensation etc. If the liability model is useful, it is not as a general account of when we are morally required to mitigate harm, but rather as a paradigm for assigning such responsibility within a particular institutional context: namely, within the law of torts. This area of law is concerned exclusively with cases in which one party (the plaintiff) alleges that another (the defendant) has caused them a loss through culpable behaviour (other than a breach of contract). Given this heavily restricted context, a court can establish that the defendant owes the plaintiff damages on the basis that they have, as alleged, culpably caused them harm. However, we should not take this to imply that the liability model is valid as a general theory of when we are obligated to mitigate harm; nor should we take those who use this model as a paradigm for assigning tort liability to be making such a claim.

But even if the social connection model were tenable (we have now arrived at my second criticism), it is unclear what implications this would have for the theory presented in the preceding chapters. In Chapter 1 (Section 2.2), I offered an interpretation of Rawls' claim that the principles of justice apply primarily to the basic structure of society. Read in the way I propose, Rawls' view is that these principles require us to shape the basic structure so that it better conforms to their specifications. Notice that this view does not define individuals' responsibilities for social justice in terms of either the liability or the social connection model. Individuals do not, that is, bear responsibilities to mitigate injustice by virtue of playing a causal role in bringing it about: either culpably or non-culpably. Rather, members of a society are required to help make their society more just purely by virtue of the fact that they are members of it.[64] This is not to say that Rawls' theory has no space for the idea

[64] In discussing Hannah Arendt's views on political responsibility, Young (2011, p. 79) describes the claim that common membership of a society can ground responsibility for justice as a "mystification". However true this may be of Arendt's position, Young's

that a person can have a moral claim to restitution if they have been harmed by another. Rather, as I explained in Section 1.1, the point is that such claims are distinct from those grounded in social justice: the former sort of claim requires the tortfeasor to repair the harm they have caused, whilst the latter requires all members of society to reform the basic structure so as to eliminate social injustice, even if they have played no causal role in its existence.

Given this, Young's argument does not show that if we introduce the notion of structural injustice into Rawls' theory, then the requirements of justice will turn out to be more extensive than those which he posits. If sound, her argument demonstrates that if we acknowledge that injustice can be structural, then this reveals the inadequacy of theories which assign responsibility for mitigating injustice using only the liability model. Introducing the social connection model into such theories, moreover, will deliver the result that our responsibilities for justice are more extensive than they initially supposed them to be. Yet since Rawls' theory is not one which defines individuals' responsibilities with respect to justice in terms of the liability model, Young's argument does not touch it. It does not, that is, provide us with any reason for thinking that the requirements which this theory makes of individuals must be extended in light of the fact that injustice can be explained structurally.

2.3. The requirements of justice: collective harm

Young's own arguments for the claim that introducing the notion of structural injustice into a theory of justice will lead it to posit more extensive requirements on individuals are unconvincing. But this need not imply that this claim is implausible. Perhaps another argument can be offered in place of Young's.

---

discussion does not touch on the principal reason which Rawlsians give for endorsing this claim: namely, the conception of society as a system of social cooperation. An adequate discussion of this idea is beyond the scope of this chapter. My point is simply to highlight this difference between Young's and Rawls' views.

A possible replacement runs as follows. Rather than thinking that justice requires things of us when we do things which partially cause injustice, we might instead think that justice requires us not to do those things in the first place. That is, one might think that we are required not perform actions which collectively produce, or reproduce, injustice. Of course, if we think this, then we may also think that those who perform such actions owe something (compensation, apology, or whatever) to the victims of injustice. But this is a further claim. It is compatible with, but not identical to, the idea currently being proposed: that a theory of justice should posit that individuals are obligated not to act in ways which collectively result in injustice.

If this claim is true, then recognising the existence of structural injustice will not give us reason to introduce any new requirements on individuals. But it will have the result that the requirements already posited will apply across a broader range of circumstances. More specifically, if we acknowledge that much injustice can be explained structurally, then we must also acknowledge that much injustice can be described as the collective result of the actions of many individuals; and if we acknowledge this, then we will arrive at the conclusion that a good deal of behaviour which we had previously thought to be anodyne is in fact morally tainted. As Young notes: "Many … circumstances that we judge unjust are … outcomes of the normal and accepted actions of millions of individuals" (Young 2011, p. 64). But if we are morally required not act in ways which collectively cause injustice, then such "normal and accepted" actions will contravene that requirement.[65]

The claim that we are obligated not to do things which collectively produce, or reproduce, injustice is endorsed by a number of moral and political theorists, many of them either influenced by or influencers of Young. For instance, Andrea Sangiovanni (2018) explicitly claims that we are morally required not to contribute causally to social processes which sustain structural injustice. Under certain conditions (such as non-culpable ignorance), he

[65] This is not necessarily to say that they will be morally impermissible. That will depend on the countervailing considerations present in a given situation, as well as one's view of moral dilemmas.

claims, it may be excusable for someone to contribute to such a process. However, whilst we may not always be at fault for doing so, we are nonetheless obligated not to do it.

Similarly, Emily McTernan's (2018) account of the wrong of microaggressions relies on the assumption that it is wrong to contribute causally to injustice. Microaggressions, for McTernan, are "subtle, innocuous, preconscious, or unconscious degradations, and putdowns" (Pierce 1995, p. 281, quoted in McTernan 2018, p. 263). Examples include: men telling a woman to "cheer up" whenever she does not present herself to them with a show-girl smile; a white woman clutching her purse to her chest whenever she passes a black man on the street; or white individuals complementing native English speakers of Asian descent on their English language abilities (McTernan 2018, p. 263). Whilst microaggressions vary in both their form (they could be expressed verbally, or through body language) and content (jokes, insults, well-intended but invalidating compliments), McTernan argues that they are unified by the fact that, collectively, they form social practices which sustain unjust hierarchies. Furthermore, she claims that this unifying feature explains what is wrong with engaging in microaggressive behaviour.

And again, Christopher Kutz writes:

> We are – at least in our more reflective moments – often troubled by our participation in collective harms, in spite of our superfluous role in producing them. This inchoate sense of unease can, I think, be seen as having two components: a sense that we, as individuals, do wrong in perpetuating the harm, and a sense of accountability towards those who suffer from it (Kutz 2000, p. 176)

By contributing to harmful behavioural patterns, Kutz thinks, we act wrongfully.[66] Moreover, Kutz's examples (environmental damage; US gun culture, and its connection with violent crime) suggest that he intends this conclusion to apply to behaviour which contributes causally to the sorts of social injustice with which Young is concerned.[67]

But why should we think this? What grounds do we have for thinking that we are morally required to refrain from actions which, together with others, collectively bring about, or sustain, injustice? As some of the above examples show, this claim is frequently presented as a truism which can be assumed, or with which a moral theory must concur. It is rarely presented as something which requires defending. This, I suggest, is a mistake.

The claim that we have a reason not to contribute causally to injustice can, I suggest, be construed as an application of a more general normative principle: namely, the claim that one has a reason to refrain from an action, if it is one of many which will collectively cause harm. The validity of this claim will be the subject of the proceeding two chapters. Whilst it has many advocates, I will argue that this claim, the 'collective harm premise', requires defence. This is because it leads to a paradox; that is, when combined with other seemingly plausible claims, it entails a contradiction. In Chapter 5, I will argue that the most plausible way in which to resolve this paradox is to reject the collective harm premise. As such, the combined argument of Chapters 4 and 5 will be that one does not have a reason to refrain from an action purely in virtue of the fact that it is one of many which will collectively cause harm.

In the context of the present discussion, this result might seem particularly troubling. How are we to break out of the causal processes which reproduce structural injustice, if individuals have no reason to refrain from collectively harmful actions? However, the view I will argue for in Chapter 5 does not imply that we could not have such a reason; indeed, it implies that if

[66] For further discussion of Kutz's view, see Chapter 4, Section 3.2.2.
[67] Indeed, Young (2011, p. 103) directly refers to Kutz's discussion of environmental damage, claiming that it is an example of structural injustice.

individuals do not have a reason to refrain from certain collectively harmful actions, then (other things being equal) they ought to have one. More specifically, I will argue that lawgivers are morally obligated to enact legislation which gives people reasons not to engage in collectively harmful behaviour.

Of course, there will be many cases in which it will be inappropriate for legislators to proscribe behaviour which collectively sustains injustice. For instance, consider the example of the gendered division of housework discussed in Chapter 1: following Cohen's lead, I posited that the fact that women tend to do more housework than men helps to sustain, in complex ways, unjust gender inequality. Does this mean that lawmakers ought to introduce legislation which proscribes heterosexual women from doing housework, or which directs their partners to do more? An affirmative answer would seem incautious, to put it mildly. But does my view therefore conclude that no one ought to do anything to stem the causal processes behind gender inequality? It does not. In cases where an outright prohibition on collectively harmful behaviour would, for whatever reason, be inappropriate, lawgivers will be obligated to use their legislative powers either to mitigate or to pre-empt the harmful effects of such behaviour by other means. For instance, various legislative provisions (gender-neutral parental leave, subsidised childcare, divorce law reforms etc.) may leave couples free to divide up housework as they choose, whilst protecting women from the potential harms of bearing the bulk of domestic responsibilities.[68]

2.4. The basic structure

What, then, is the philosophical significance of the concept of structural injustice? Not that it reveals injustice which would otherwise be obfuscated. Not that it forces us to posit additional individual responsibilities for mitigating injustice. And not that it uncovers the moral taint of much putatively anodyne, everyday behaviour. So, does acknowledging that injustice can be structural contribute anything to a theory of justice? Despite

[68] See e.g. (Okin 2005; 1989).

the arguments of the preceding subsections, I think that it can. The recognition that injustice can be explained structurally, I submit, strengthens the case for thinking that the principles of justice must apply primarily to the basic structure of society. This result is suggested by various points in the discussion of the preceding subsections.

Consider, for instance, Haslanger's observation which I quoted towards the end of Section 2.1: the structural constraints which individuals face can make it rational for them to make decisions which lead them to suffer injustice, or which compound injustice which already confronts them. Such situations can be seen as one indirect way in which the basic structure (society's major institutions considered collectively) can determine the societal distribution of fundamental rights, duties, and benefits: the major institutions (including the family) determine the principal positions or offices which individuals can occupy, and these in turn determine choice-situations which make rational certain choices that collectively result in a given distribution. Given this, an awareness of the ways in which structural constraints condition people's choices will be a valuable resource for a theorist of justice. Such an awareness will reveal the pervasiveness of the influence of society's major institutions on people's life prospects; these institutions effect their influence not merely directly, by means of their explicit systems of rules, but also indirectly, by way of defining the principal social positions which people can occupy. Thus, if our aim is to achieve and sustain a fair distribution of rights, duties, and benefits, then this will be even more reason to concentrate our efforts on shaping the basic structure.

A further consideration in favour of my conclusion can be gleaned from the discussion of Section 2.3. There I pointed out that if we acknowledge that much injustice can be explained structurally, then we must also acknowledge that much everyday behaviour contributes to largescale behavioural patterns which collectively produce and reproduce injustice. However, pre-empting the arguments of the proceeding chapters, I claimed that this does not necessarily show us that such behaviour is morally objectionable: since the fact that an action is one of many which will collectively cause harm does

not, by itself, give an individual a reason to refrain from it. Rather, what this shows us is that lawgivers are obligated to enact legislation which will give individuals reasons to avoid behaviour that collectively sustains injustice, or at least which will mitigate or pre-empt the harmful effects of such behaviour.

Legislative acts exercise a particularly strong influence over the basic structure of society; such acts not only alter a society's legal system but can also effect changes in other institutions: either by instigating changes to their explicit rules, or by prompting changes to informal, conventional behaviour. As such, the obligation on lawmakers to introduce legislation which protects individuals from structural injustice can be seen as an instance of the wider obligation to shape the basic structure in accordance with the stipulations of the principles of justice. In other words, the specific requirement for lawmakers to protect people from injustice-sustaining behavioural patterns can be derived from the individual requirements implied by the principles of justice. This, I submit, gives us all the more reason to retain Rawls' claim that the principles of justice apply to the basic structure. If the most plausible response to the existence of structural injustice is to think that legislators are obligated to protect us against the unjust, aggregate effects of largescale behavioural patterns, and Rawls' claim already implies this response, then so much the better for Rawls' theory.

Appendix: Injustice as harm

At various points in this chapter, I have relied on the assumption that injustice is a form of harm. I made this assumption, for instance, when discussing Young's views on responsibility for injustice: Young presents two models of responsibility, both of which are supposed to assign responsibility for mitigating injustice on the assumption that it constitutes harm. It was also assumed in my discussion of collective harm: I suggested that we could understand the claim that we have a reason not to contribute causally to injustice as an instance of the more general claim that we have a reason not to perform collectively harmful acts. But is this assumption plausible? Is it plausible, that is, to think of social injustice as a kind of harm?

In the first instance, it might seem strange to think of social injustice in this way. Unjustified inequalities in wealth, opportunities, or liberty do not seem to harm us in the same way as a stab wound, or a divorce, or an illness. Of course, if someone is so badly off that that cannot meet their basic needs, then they will likely suffer harm in the form of disease or malnutrition. Or if a person's basic liberties are severely restricted by some oppressive regime, then they may suffer physical violence at the hands of government thugs. But the victims of social injustice need not suffer illness or injury.

To illustrate, imagine that a government implements a 'separate but equal' policy, under which individuals from one ethnic group are banned from associating with members of another ethnic group. Individuals of both groups are, however, ensured the same liberties and the same kinds of opportunity. If an individual from either group wants to be, say, a doctor, or a lawyer, or an athlete, then she can; she just can't train or work alongside members of the other group. Moreover, the jobs open to one ethnicity pay just as well as those open to the other. And if an individual is unable to secure salaried work, then, regardless of her ethnicity, she is entitled to a generous welfare stipend.

This imagined society is unjust: its citizens are denied freedom of assembly and association, a freedom which Rawls lists among the basic liberties to be guaranteed under the liberty principle (Rawls 1999, p. 53). And yet it is not immediately clear that anyone is harmed by this unjust policy. It does not cause anyone to suffer absolute poverty, and the illnesses associated with that condition. Nor, let us stipulate, does anyone suffer physical violence at the hands of the policy's enforcers. Relations between the two ethnic groups are peaceful but cold, and as such there is no need for harsh penalties to incentivise compliance.

Despite what such examples seem to suggest, however, the idea that social injustice is a form of harm seems to me defensible. This claim gains plausibility, I suggest, if we conceive of harm as the setting back of a person's interests. Such a conception of harm lurks in the background of a variety of

discussions.[69] It receives explicit and extended discussion, however, from Joel Feinberg (1984, chap. 1). For Feinberg, the notion of harm admits of a variety of different interpretations. On one of its central senses, however, a person is harmed just in case they face a setback to one or other of their interests, where a person's interests are the various distinguishable components of their wellbeing, the things which make their life go better or worse. These range from the satiation of their basic biological needs (*welfare interests*, in Feinberg's terminology), to the achievement of more ultimate, comprehensive goals (*ulterior interests*) (Feinberg 1984, p. 37).[70]

Clearly, Feinberg's account implies that a person suffers harm when they live in absolute poverty. Under such conditions, a person cannot satisfy their own basic biological needs, and so suffers a setback in their welfare interests. Yet it also implies, I suggest, that a person is harmed whenever the principles of justice imply that they can legitimately complain about their situation. The basic rationale behind the Rawlsian principles is that they enable individuals to pursue their own comprehensive conceptions of the good. These principles are meant to ensure everyone the maximal range of social primary goods: that is, goods which it would be rational for a person to want, regardless of what else they want. As such, whenever the requirements of one of these principles goes unmet, there will likely be some individual who faces difficulties in pursuing her conception of the good as a result; at least someone will be lacking some social primary good which would better enable her better to carry out her comprehensive goals. Yet to be in such a condition is, according to Feinberg, to be harmed; a person who is obstructed from following her ultimate projects and aspirations faces a setback in her ulterior interests. As such, social injustice implies harm.

To see this more clearly, let's return to the 'separate but equal' policy imagined a few paragraphs ago. I said that it was unclear how this policy harmed anyone, since it did not cause anyone to suffer illness of any sort, nor

[69] Notably, in exegetical discussions of Mill's harm principle. See e.g. (Rees 1960; Waldron 1987).
[70] For a critical discussion of Feinberg's account, see (Shiffrin 2012).

was anyone injured by the policy's enforcers. Assuming Feinberg's conception of harm as setbacks to interests, we can now see the harm which this policy inflicts. The policy denies individuals freedom of association. This freedom is among the basic liberties required for individuals to pursue various conceptions of the good. As such, to deny people freedom of association is to obstruct them from advancing a range of ulterior interests.

One might object that the comprehensive goals of the members of this particular society are not obstructed by the 'separate but equal' policy. I stipulated that relations between the two segregated ethnic groups are cold. As such, one might infer that the members of these groups have no interest in associating with one another; through being socialised under segregation, these individuals have adopted conceptions of the good which require them not to associate with members of the other group. But if this is so, then which of these people's interests are impeded by denying them freedom of association?

I grant that if individuals hold segregationist conceptions of the good, then the 'separate but equal' policy will not harm them. However, it does not follow that no one is harmed here. Whilst the policy itself may not cause harm, the fact that individuals hold such conceptions of the good, I suggest, does. It is difficult to see what could ground a stable desire not to associate with members of another ethnicity other than plain racism.[71] Yet the fact that people hold racist attitudes towards another ethnic group can be seen as a setback to the interests of the members of that group. Individuals have interests in the way in which others represent them: not least because a person's conception of herself, her value and powers, is intimately tied up with others' representations of her.[72] Racism, of the sort currently under consideration, involves the systematic misrepresentation of another ethnicity.

[71] By "plain" racism I mean, roughly, what Tommie Shelby calls "ideological" racism: "a widely held set of associated beliefs and implicit judgements that misrepresent significant social realities and that function, through this distortion, to bring about or perpetuate unjust social relations" (Shelby 2016, p. 22).
[72] See for instance Rawls' discussion of self-respect (Rawls 1999, pp. 386-391).

As such, the existence of such racism, even if purely in the mind of the racists, can be construed as harmful.

At this point, my objector might try a slightly different tack. They might point out that if members of one ethnicity tend to avoid members of another, it does not follow that they all must endorse a racist conception of the good. For instance, whilst their comprehensive values and ideals might be entirely consistent with their befriending and loving members of other ethnic groups, the segregated nature of their social environment might mean that they are never presented with the opportunity to do so. Moreover, one might think that the imposition of the 'separate but equal' policy does not harm such individuals, since it does not obstruct the activities which they in fact choose to pursue. Yet neither do their conceptions of the good cause harm, as they are not racist. So, if our imagined segregated society is one in which people comply with the 'separate but equal' policy not because they endorse racist ideology, but rather because the circumstances in which they would transgress it never arise, then it seems we will have injustice without harm.

It is surely correct that compliance with the 'separate but equal' policy need not be fuelled by overt racism. Nevertheless, there seems room for doubt about the claim that imposing this policy upon individuals whose compliance with it is not fuelled by racism does no harm to them. For any conception of the good (or least, any sufficiently complex one), there will be various possible ways in which an individual may pursue it: there will, that is, be a range of possible courses of action the pursuit of which will be consistent with an individual's endorsing that conception. If sufficiently many of these possibilities are blocked off that one is impelled to live out only a few of the full range given by one's conception of the good, then it seems not unreasonable to think that one is impeded in pursuing that conception. As such, whilst denying freedom of association to those in the segregated society may not frustrate anyone's actual plans, it may nevertheless inhibit their pursuit of their conceptions of the good; if forbidding members of different ethnicities from mixing constrains the range of activities which they can follow in pursuit those conceptions to an intolerably small subset, then it is

not so implausible to think that it frustrates their ulterior interests. As such, the ‘separate but equal’ policy will harm these individuals.

Part II

Collective Harm

4

The Problem

Towards the close of the preceding chapter, I raised the following question: Are we morally required not to do things which collectively reproduce social injustice? Given some further assumptions (namely, that social injustice is a harm, and that having a reason not to something is a prerequisite for bring morally required not to do it), our answer to this question will depend on our answer to a more general question: If an action will be one of several which will collectively cause harm to others, do we therefore have a reason to refrain from it? This second question will be the subject of this chapter and then next. It might seem plausible to answer this question in the affirmative. However, in Section 1 of this chapter, I outline a problem which this answer encounters when we consider that several actions might together cause harm, without any one of them making any difference. Section 2 develops this problem in greater detail; and Section 3 assesses five potential solutions which have been put forward in the literature. One solution of particular interest is that proposed by Julia Nefsky. Nefsky's account has gone largely unchallenged. In the course of Section 3, I probe a possible weak point in the view.

1. A tension

Many of today's most pressing moral and political issues concern cases in which harm is caused not by what any one person has done, but rather by the actions of many individuals. Concrete examples abound: the climate crisis, for instance, is, in large part, the cumulative result of the ways in which millions of individuals worldwide consume energy. Chapter 3 offered a more generalised illustration of the same point: various instances of social injustice can be explained structurally, where this involves explaining a given social condition as the aggregate effect of many individuals' rational decisions.

Reflecting on this observation can make us uneasy; it can be troubling to think that one's actions form a part of behavioural patterns which collectively sustain harmful social conditions. It is quite natural to suppose that we at least have some reason, perhaps even a moral obligation, not to do things which together cause harm to others.

However, whilst natural, this thought faces a problem. It seems possible that several actions might together cause a harmful outcome, whilst no one of them makes any difference to that outcome. To keep things simple, consider a toy example:

The Reservoir: A large number of individuals (the tippers) each possess equal amounts of industrial waste of which they wish to dispose. When consumed in small quantities, the waste is completely harmless. However, if consumed in larger quantities, it is toxic. A potential dumping ground is a local reservoir, on the banks of which lies a hamlet whose inhabitants use the reservoir as their main source of drinking water. The quantity of waste possessed by each of the tippers is so small that, regardless of how much waste is already in the reservoir, if any one of them were to add their share to its waters, it would make absolutely no difference to the villagers' health. However, if all, or sufficiently many, of the tippers were to do so, then the villagers would be poisoned as a joint result.[73]

[73] This example differs from those typically considered in the literature. Often, discussions focus on the possibility that several actions might be collectively painful, yet severally painless. This possibility is supposed to arise when the actions of many individuals together cause someone to suffer pain, whilst the effect of each one of those actions considered singly is imperceptible to the victim; see, for instance, Derek Parfit's well-known 'Harmless Torturers' example (Parfit 1984, p. 80). It seems to me, however, that a focus on such cases introduces unnecessary noise. First, it is unclear whether we can infer that an action causes a person no pain from the fact that they cannot perceive its effects; see, for instance, (Hedden 2020; Parfit 1984, pp. 78-82). Second, on many accounts of harm, not all instances of pain are considered harms; see (Shiffrin 2012) for discussion of this point. But if we grant this, it becomes less obvious that cases of collectively painful/individually painless actions necessarily pose a problem; if the pain caused by those actions is not an instance of harm, then it is not immediately clear that people ought to refrain from them. I suggest that these two difficulties can be bypassed by concentrating on cases in which several actions seem to be both collectively harmful and severally harmless, but where the harm in question is not pain; the reservoir is supposed to be such an example. Focussing on cases like this evidently allows us to avoid the second difficulty. But it also avoids the first. Whether or not an action is harmful seems to be entirely independent of whether or not its effects are perceptible. To illustrate, suppose that one of the villagers possesses equipment which enables her to perceive

When faced with such cases, our intuitions seem to pull us in conflicting directions. On the one hand, since these actions will together harm the villagers, we want to say that the tippers should refrain from tipping their waste in the reservoir. But on the other, because each waste-tipping makes no difference, it is difficult to see what reason they have to refrain. To put the point differently, it seems that each tipper could acknowledge that it would be a bad thing if the villagers were to be poisoned, and yet still doubt that they have any reason not to tip their waste into the reservoir: 'Afterall, *my* waste-tipping won't cause anyone any harm'. This sort of response can seem both pertinent and evasive: it seems true that, considered individually, each act will be harmless; but it is also true that, taken together, those acts will cause harm. When we reflect on the 'but-I-make-no-difference' retort, our thoughts seem to oscillate between these two points.

This tension has been the subject of much philosophical discussion over the last few decades.[74] In the following chapter, I will offer my own thoughts on how best to resolve it. In this chapter, however, my focus will be restricted to two prior questions: First, what exactly is the problem posed by cases like that of the reservoir? And second, are the existing solutions to that problem tenable?

2. A paradox

In the foregoing paragraphs, I formulated the difficulty posed by the possibility of cases like the reservoir in terms of a tension between two intuitive thoughts: since the tippers' actions will together cause harm, we want to say that they have a reason to refrain; and yet, since each waste-tipping will be individually harmless, we are also tempted to say each tipper

very minute levels of waste in the reservoir's water. Does this mean that she will be harmed if she consumes such tiny amounts of waste, whilst others who lack her perception-enhancing tools will not? It would seem odd to think so. See (Nefsky 2011, pp. 373-375) for a similar point.

[74] Current discussions of the problem were sparked largely by Jonathan Glover (1975) and Derek Parfit (1984, chap. 3). For some recent contributions see e.g. (Barnett 2018; Dietz 2016; Hedden 2020; Kagan 2011; Nefsky 2017; Pinkert 2015; Sangiovanni 2018; Spiekermann 2014). See (Nefsky 2019) for an extensive bibliography.

has no reason to refrain. My primary aim in this section will be to develop and motivate this problem. Yet I will also have two subsidiary aims.

First, the difficulty which I have been gesturing towards is often characterised in a way which presupposes a particular kind of solution. A few paragraphs ago, I pointed out that a tipper might attempt to defend their decision to tip their waste into the reservoir by say something like 'But my act makes no difference!', and that there is something at once compelling and unpersuasive about this kind of thought. Several authors suppose that the challenge posed by these sorts of case is to show where that thought goes wrong. For instance, Julia Nefsky writes:

> The problem in such cases is that if acting in the relevant way won't make a difference, it's unclear why it would be wrong. Each individual can argue, "things will be just as bad whether or not I act in this way, so there's no point in doing otherwise." It seems there must be something wrong with this argument. After all, when enough people reason in this way, serious avoidable harm results from our voluntary actions (Nefsky 2019, p. 2)

What we need to do, thinks Nefsky, is demonstrate the falsity of the claim that the tippers, and others like them, have no reason to refrain from their collectively harmful actions; that claim must be false, the difficulty is proving it. However, to formulate the problem in this way is to approach it with a prior conception of how it should be resolved. What we have, I suggest, are two pretheoretically attractive claims which, it seems, cannot both be true: very roughly, the claim that we have no reason to avoid individually harmless acts; and the claim that we have a reason to avoid collectively harmful acts. Why assume, at the outset, that it is the first of these two claims which must go? Why not leave it open? In this section, I aim to characterise the problem at

hand in a way which does not make assumptions about what its solution should look like.[75]

Second, it is typically supposed that cases like the reservoir pose a problem for a particular kind of moral theory: namely, consequentialism.[76] It is easy to see why discussion of such cases focusses on this theory. Consequentialism (or, at least, *act*-consequentialism) is the view that an action's permissibility is determined entirely by how good its consequences are. In its crudest form, the theory says that an action is morally right if and only if, out of the possible actions available to the agent at that time, it produces the best consequences; otherwise, it is morally wrong. The reservoir seems to be precisely the kind of case of which a consequentialist would want to say that the actions which stand to be performed are morally wrong; it is precisely the nasty cumulative consequence of the waste-tippings which seems to worry us here. However, consequentialism seems unable to deliver this result. If we restrict each tipper's options to {tip, refrain}, and restrict the relevant outcome to the quality of the villagers' health, then it seems that the possible actions facing each tipper will produce equally good outcomes: since, holding others' actions fixed, the villagers' health will be precisely the same whichever option they pick. But if so, then consequentialism cannot give us the result that tipping would be morally wrong. As Shelly Kagan puts it: "if there are indeed case that have this sort of structure … then consequentialism appears to fail even in its own favored terrain, where we are concerned with consequences and nothing but consequences" (Kagan 2011, p. 107).

However, whilst I do not deny that cases like the reservoir seem to pose difficulties for consequentialism, I want to resist the trend of focussing on these particular difficulties. The possibility of such cases, I suggest, does not only create problems for people who happen to think that that ethical theory

[75] Nefsky does address the view that we should deny that we have a reason to refrain from collectively harmful acts. However, she claims that such views are sceptical, since they deny that the problem has a solution (Nefsky 2019, pp. 2-3). This, I am suggesting, is to presuppose that views of this sort do not constitute a solution.
[76] See e.g. (Glover 1975; Hedden 2020; Kagan 2011; Parfit 1984, chap. 3; Pinkert 2015). Nefsky (2015) is a notable exception.

is true. It generates a tension between two claims which people who reject consequentialism can nonetheless find attractive. I will return to this point in Section 2.4. For now, however, I simply want to note that my intended audience is broader than the class of card-carrying consequentialists.

Let's turn, then, to the problem at hand. I suggest that the difficulty posed by cases like the reservoir can be formulated more precisely as a paradox: that is, as a set of initially plausible, but inconsistent, claims:

The paradox of collective harm

(i) ***The collective harm premise:*** If an action is one of several which will collectively cause harm to others, then one must have a reason not to perform it.

(ii) ***The inefficacy premise:*** One has no outcome-reason to refrain from an action if, considered by itself, it is harmless.

(iii) ***The harmful/harmless series premise:*** It is possible:

(a) for a series of actions to be collectively harmful, whilst each member of that series is individually harmless; and

(b) for there to be such a series such that if the relevant individuals have a reason to refrain from their collectively harmful actions, then it must be an outcome-reason.

Premise (iii) is intended to capture the structure of cases like that of the reservoir, whilst premises (i) and (ii) are meant to express the two sorts of intuitive reaction which cases of that type evoke. Yet, taken together, these three claims entail a contradiction:

(iv) An individual can both have a reason and have no reason to refrain from performing one of a series of collectively harmful/individually harmless actions.

The fact that a set of claims entail a contradiction is not, however, sufficient to render them a paradox; {one plus one is two, one plus one is three} is not a paradox. In addition, each of those claims, considered on their own, must

have some plausibility. The proceeding subsections take up each of the three premises of my alleged paradox in turn, clarify their meaning, and showcase their independent attraction.

2.1. The collective harm premise

Let's start with the collective harm premise: if an action is one of several which will collectively cause harm to others, then one must have a reason not to perform it. Variants of this claim are very widely endorsed in the literature on the ethics of collective action.[77] However, few authors have paused to investigate the grounds we have for supposing it to be correct. Moreover, the paradox of collective harm demonstrates that its truth cannot be safely assumed; so long as each member of the triad retains its independent appeal, all are suspects.

Before considering possible grounds for the collective harm premise, however, it will be helpful to distinguish it from a similar, though implausible, claim:

The collective harm counterfactual: The fact that many actions of a given type would, if enacted together, cause harm to others provides a reason not to perform an action of that type.

This claim is demonstrably false. For example, it would be disastrous if many people were to crush into a tightly enclosed space. Do I thereby have a reason not a enter a very small, but empty, passageway? No. However, this implausible principle is not equivalent to, nor does it follow from, the collective harm premise. Unlike the collective harm counterfactual, the

[77] The collective harm premise seems to be implied, for instance, by Kai Spiekermann's principle "Contribution is wrong: Any action that contributes to the impact is wrong" (Spiekermann 2014, p. 78), where an action contributes to an impact if it makes a difference to some precisely measurable factor which is causally related to a harm (for instance, the level of waste present in the reservoir). It is also very close to Felix Pinkert's "On-the-hook: In any collection of agents who together gratuitously fail to bring about collectively optimal outcomes, there must be some relevant morally objectionable facts about some of the agents" (Pinkert 2015, p. 975). Both Spiekermann and Pinkert intend for their principles to distil similar claims made by various contributors to debates on the ethics of collective action.

collective harm premise does not imply that one has a reason not to φ, if many acts of φ-ing would jointly cause harm. Rather, it states that one has a reason not to φ, if one's act of φ-ing would be one of several actions which *will in fact* collectively cause harm. As such, it does not imply that we have a reason to avoid entering tightly enclosed, yet empty, spaces. Rather, more plausibly, it entails that an individual has a reason not to enter such a space, if that action would be one of several which will together cause a crush.

Having cleared one possible cause for doubt, let's turn to consider some possible grounds for endorsing the collective harm premise. One might propose that we should endorse it simply on the basis of its intuitive appeal; when we reflect on this principle, we 'see' that it is correct, and so we should endorse it on that basis. Such appeals to moral intuition cannot be dismissed out of hand. Yet, as Rawls (1999, pp. 36-37) points out, it is preferable if they can be avoided. When a principle is endorsed purely on the basis of intuition, there will no way of deciding cases in which it conflicts with other intuitive principles beyond further appeal to intuition. There is nothing irrational in this. Yet it is to admit an end to moral argument; when individuals disagree about how to resolve a conflict between two intuitive principles, there will be nothing more for them to say.[78]

So, if we can provide reasoned grounds for adopting the collective harm premise, rather than simply appealing to intuition, then it will be preferable to do so. Can such grounds be found? I think they can. I propose that the collective harm premise can be defended on abductive grounds. To see this, suppose that all of the tippers in the reservoir case chose to dump their waste into the water, with the foreseeable result that the villagers were poisoned. The harm caused in this case would not, it seems, strike us as merely unfortunate, or regrettable. Rather, it would strike us as *morally objectionable*. Unlike that caused by a non-culpable accident, or an 'act of God', the harm which the villagers stand to suffer is not something we think

[78] "The assignment of weights is an essential and not a minor part of a conception of justice. If we cannot explain how these weights are to be determined by reasonable ethical criteria, the means of rational discussion have come to an end" (Rawls 1999, p. 37).

it would be acceptable to ask them to bear; these individuals, it seems, have a valid moral claim against being harmed in this way. This, I take it, is the force of Nefsky's remark which I quoted at the outset of this section: "After all, when enough people reason in this way, serious avoidable harm results from our voluntary actions" (Nefsky 2019, p. 2). The relevant point, however, is not that the harm is serious or avoidable *per se*, but rather that it is objectionable.

To make this more apparent, suppose for a moment that I possess the same amount of waste as that possessed collectively by all of the tippers. If I were to dump this, all in one go, into the reservoir, then the harm suffered as a foreseeable result of my voluntary action would surely be objectionable. But if we think this, then shouldn't we also think that it would be objectionable for the villagers to be equally harmed as a foreseeable result of many individuals' voluntary actions? From a moral point of view, the number of individuals whose actions have caused the harm seems irrelevant.[79]

How can we explain this datum? Why is it that villagers have a valid claim against being harmed as an aggregate result of the tipper's actions? Throughout this chapter and the next, I will assume that harm is morally objectionable if at least someone failed to take morally obligatory steps to prevent it. That is, an individual has a valid claim against being harmed in some way if someone else has an obligation to take certain reasonable precautions to protect them from it.[80]

Of course, harm can be objectionable even when it does not occur because of some moral failing. For instance, if a hurricane devastates a neighbourhood, and the residents are offered no aid from their government, then the harm they suffer may be morally objectionable: even if every reasonable precaution was taken to protect the residents from the hurricane's devastation. However, this sort of observation will not help to explain our moral judgements about cases

[79] A similar point is made by Glover (1975, pp. 174-175).
[80] For an elaboration of this point, see (Scanlon 1998, p. 257).

like the reservoir. The harm caused by the hurricane is objectionable, it seems, because others owe it to the victims to help alleviate their suffering, yet fail to do so; if these others were to offer their help, then the harm would no longer seem so objectionable. Yet the objectionability of the villagers' poisoning cannot be accounted for in this way. If the tippers were to dump their waste in the reservoir, thereby poisoning the villagers, but then distribute medicines to mitigate the suffering, we would not be content that things were as they should be. The villagers, it seems, should not have been poisoned in the first place.

Given all of this, it seems very plausible to explain the objectionability of the villagers' being poisoned by supposing that the tippers are morally obligated not to dump their waste into the reservoir. If the tippers were obligated not to tip, then any harm caused collectively by their waste-tippings would be morally objectionable. These individuals would be obligated to take steps to prevent the villagers from being poisoned: namely, by refraining from tipping their waste into the latter's drinking water. As such, the villagers would have a claim against being harmed in that way.

But if we adopt this explanation, then we seem to be committed to the truth of the collective harm premise. Acts which are morally impermissible are acts the performance of which cannot be justified to others; and if we cannot justify a given type of act to others, this is because there is at least some reason not to perform it which is sufficient to countervail other reasons which might count in its favour.[81] So, if the tippers are obligated not to tip, then they must have a reason not to do so. But what could this reason be, if it is not the fact (or, at least, implied by the fact) that these actions together cause harm to other people? There seems to be no other feature of the case which could count against the tippers' actions. This, I propose, provides an abductive case for the collective harm premise: harm caused collectively by many actions

[81] This is to make of impermissible acts the converse of the point which Scanlon (1998, p. 197) makes of justifiable acts: "To justify an action to others is to offer reasons supporting it and to claim that they are sufficient to defeat any objections that others may have." For further discussion of a similar point see Chapter 1, Section 2.2.

can, we think, be morally objectionable; and this is best explained by appealing to the collective harm premise.

Before moving on, I want briefly to note a potential exception from the collective harm premise. Suppose, for a moment, that we modify the reservoir case. In this modified example, I stand alone at the water's edge, holding the same miniscule quantity of waste as that possessed by each one of the tippers in the original example. However, a factory which has been built on the banks of the reservoir regularly dumps huge quantities of waste into the water; indeed, it dumps so much that the villagers' poisoning is already guaranteed.

If I were to add my miniscule quantity to the reservoir's waters, then my action would be one of several which collectively cause harm; the villager's would be poisoned because of the amount of waste present in the reservoir, and this in turn would be the cumulative result mine and the relevant factory employees' actions. Do I therefore have a reason not to dump my waste into the reservoir?

Intuitive reactions may vary. Some might think that, since it is settled in advance that the villagers will be poisoned, and my action is certain to make no difference, I have no reason to abstain. Indeed, some solutions to the paradox of collective harm, such as Nefsky's and Spiekermann's, entail this result (see Section 3.4).[82] Others, however, might disagree; regardless of whether or not the harm is guaranteed in advance, one always has a reason not to implicate oneself causally in the production of harm. For the purposes of this and the following chapter, it does not matter which of these reactions one endorses. As such, I will leave both options open.

2.2. The inefficacy premise

Let's turn to premise (ii) of my alleged paradox: the inefficacy premise. This premise, recall, says that an individual has no outcome-reason to refrain from an action if, considered singly, it is harmless. Attentive readers will notice

[82] Spiekermann (2014, p. 89) explicitly endorses this implication of his view.

that I have introduced a new phrase: 'outcome-reason'. What does this new phrase mean, and why have I introduced it? Why not simply say, as in fact I did at the outset of this section, that one has no *reason* to refrain from singly harmless acts?

I will take the second question first. The reason why I have moved away from saying baldly that one has no reason to refrain from singly harmless acts is that this claim is plainly false; an individual might have a variety of reasons not to perform an action which, considered on its own, causes no harm.

An answer to the first question will help to demonstrate this. We can distinguish between two sorts of reason for action: *outcome-* and *action-reasons*.[83] An outcome-reason is a reason to act which is based on the value of the states of affairs which that action brings about, where a state of affairs is, roughly, a possible way the world might be. For instance, if I have a reason to donate to a food bank, this appears to be because my doing so bring about something good: namely, it will sate hunger. I would have no reason to do this, it seems, if it turned out that my donation would not have this consequence. An action-reason, by contrast, is a reason which is based on the value of the act it counts in favour of, rather than on that of the further consequences which a given action of that type produces. For example, the reason one has to keep one's promises does not seem to be based on the goodness of the consequences which one might bring about by keeping a particular promise. If things would go better were I to break my promise, I would not thereby lose my reason to keep it (though that reason might be outweighed).[84]

Even if an action were harmless, it seems that one could nonetheless have an action-reason to refrain from it; one might have promised, for instance. However, it is much less evident that someone could have an outcome-reason to refrain from a harmless action. One could perhaps defend that claim if one

[83] I draw this distinction from Joseph Raz (1986, pp. 145-146). Whilst it is supposed to be exclusive, it is not intended to be exhaustive.
[84] This point is elaborated by Philippa Foot (2001, pp. 47-48).

adopted a very limited notion of harm, on which not all of an action's bad consequences were necessarily harmful. However, on a broader conception of harm, it seems less defensible. It is the broader notion of harm which I will be concerned with here.[85]

It will aid the discussion later on to consider another sort of case in which one might have a reason to refrain from a singly harmless action: namely, cases of risk. We cannot always know in advance what the outcomes of our actions will be. In such cases of uncertainty, it is intuitive to think that individuals have a reason to avoid actions which pose an unacceptable risk of harm to others. Moreover, if an individual chooses to perform an unacceptably risky action which, luckily, turns out to be harmless, it nonetheless seems intuitive to think, in retrospect, that they acted unreasonably. For instance, suppose that someone drives home drunk from a party. Luckily, however, he arrives home safely. Does the harmlessness of his action imply that he had no reason to refrain from it? It seems not. The drunk driver should not have driven home because doing so posed an unacceptable risk of harm to others.

For the purposes of the present discussion, I will assume that such risk-based reasons are distinct from outcome-reasons. The lucky drunk driver's reason to refrain from driving is not grounded in badness of the actual consequences of his action: these consequences were perfectly happy. Rather, his reason is grounded in the badness of the possible consequences which his action *might* have brought about. The strength of this reason is a function of the badness of these possible consequences and the probability of the action's bringing them about. I don't pretend that this is the only way of characterising risk-based reasons, and none of the important parts of the proceeding discussion

[85] In note 73, I pointed out that there are accounts of harm which hold that some instances of pain are not instances of harm. If this is correct, then perhaps one could have an outcome-reason to avoid a harmless action if it causes non-harmful pain. However, even if we grant this exception to the inefficacy premise, it will not support the claim that the tippers have a reason to refrain from tipping their waste into the reservoir: since none of these actions cause pain.

will hang on it. However, assuming this characterisation will help to simplify things.[86]

2.3. The harmful/harmless series premise

Let's move now to the final premise of my paradox, the harmful/harmless series premise. This claim, recall, is split into two clauses. Clause (a) stipulates that it is possible for a series of actions to be collectively harmful, whilst each member of that series is individually harmless; and clause (b) stipulates that it is possible for there to be a series of this kind such that if the relevant individuals have a reason to refrain from their collectively harmful actions, then it must be an outcome-reason. Let's consider these two clauses in turn.

To begin, let me clarify some of the terms used in clause (a). For any plurality *xx*, in describing *xx* as 'collectively', or 'jointly', or 'together' *F*, I am attributing the property of being *F* to the plurality. Thus, in the sentence '*xx* are collectively *F*', the predicate *F* takes several objects as its subjects, rather than any single object. By contrast, if I describe *xx* as 'individually', or 'singly', or 'severally' *F*, then I am attributing the property of being *F* to every *x* which is one of *xx*. Thus, in the sentence '*xx* are individually *F*', *F* takes a single object as its subject: namely, any *x* amongst *xx*. In other words, '*xx* are individually *F*' can be paraphrased as a conjunction in which every occurrence of *F* is predicated of a name which denotes only one thing. For example, if *xx* = *a*, *b*, then '*xx* are individually *F*' is logically equivalent to '*a* is *F* and *b* is *F*'. By contrast, '*xx* are collectively *F*' cannot be paraphrased in this way; its meaning cannot be conveyed by any sentence in which *F* is only ever predicated of a term denoting a single object. A final piece of terminology. Whenever several things are individually *F*, they can be described as *distributively F*. By comparison, if a plurality are individually not-*F* (i.e. if no one of them is *F*), but they are collectively *F*, then the plurality

[86] Claire Finkelstein (2003) has argued that exposure to a risk of harm itself constitutes harm. If she is correct, then there will be a case for thinking that reasons to avoid risky actions can be classified as outcome-reasons. I have no objection to this way of thinking about risk, nor do I think that it scuppers the arguments of this chapter. However, the assumption that risk-based reasons are not outcome-reasons simplifies some of these arguments.

can be described as *non-distributively F*. Clause (a) of the harmful/harmless series premise claims that it is possible for a series of actions to be non-distributively harmful. That is, it claims that there is some possible series of actions *xx* such that *xx* together instantiate the property of being harmful, whilst no one of *xx* instantiates that property.[87]

The reservoir is supposed to be an example of the kind of case which clause (a) claims to be possible. The reasoning behind this supposition runs as follows. Surely, one might think, there is a unit of measurement so small that, for any number x, water containing x units of waste will be just as harmful to drink as water containing x+1 units. That is, if water containing x units is perfectly safe, then water containing x+1 units will be perfectly safe as well; the extra amount is so minute that it could never make the difference between safe and unsafe drinking water.[88]

Now suppose that each tipper possesses one of these miniscule units of waste. Suppose further that there are n tippers, and that they are named t_1, …, t_n. Finally, suppose that the tippers will only dump their waste in series, so that, for instance, t_2 will tip only once t_1 has tipped, t_3 will tip only once t_2 has tipped, and so on. The remarks made in the preceding paragraph imply the following claim. For any number $x \in \{1, \ldots, n\}$, if t_x dumps their waste, then the reservoir's water will be just as harmful to drink as it was when t_{x-1} dumped their waste (where t_0 is interpreted as 'no one').

This establishes that the series of waste-tippings are individually harmless; whilst each of these actions causes an increase in the amount of waste present

[87] The notions of collective, individual, distributive, and non-distributive property ascription are drawn from Thomas Smith (2009; 2011), and these themselves are modified from concepts in plural logic. In classical Frege-Russell calculus, any given quantifier or predicate operates over a term denoting a single object. By contrast, in a plural logic, quantifiers and predicates can operate over terms denoting a plurality of objects. Thus, whilst Frege-Russell calculus can say certain things about pluralities only by compounding them into a single object, such as a set, plural logics enable us to talk about pluralities in a more parsimonious way. See e.g. (Boolos 1984; Hossack 2000; McKay 2006; Oliver and Smiley 2013).

[88] This supposition will be challenged in Section 3.1. There, I argue that the series of waste-tippings could still be severally harmless, even if there is a point at which one more unit of waste could make the difference between safe and harmful drinking water.

in the reservoir, no single one of them causes an increase in the level of harm suffered by the villagers. However, it was stipulated that if all, or sufficiently many, of the tippers were to dump their waste in the reservoir, then the villagers will be poisoned. Therefore, clause (a) of the harmful/harmless series premise is true: it is possible for a series of actions to be collectively harmful, whilst each member of that series is individually harmless.

It will be helpful to distinguish the kind of case I have just described from another sort of case. Let's suppose now that the case of the reservoir is not as I have just described it. Instead, suppose there is some number $x \in \{1, \ldots, n\}$ such that if t_x dumps their waste, then the harm suffered by the villagers will be worse than it was when t_{x-1} dumped their waste. In this case, t_x's action 'triggers' harm to be caused; their action crosses a threshold from less harm to more. In fact, more accurately, when t_x dumps their waste, every waste-tipping preceding (and including) t_x's makes a difference to the harm suffered by the villagers; each of these actions is such that, holding other actions fixed, had it not been performed, less harm would have occurred. If the reservoir case is like this, then it will not be an example of the kind of case which the harmful/harmless series premise claims to be possible. Following Kagan (2011) and Nefsky (2011), I will call cases of this sort *triggering cases*, and cases the possibility of which would verify the harmful/harmless series premise *non-triggering cases*.

Note that triggering cases can be cases in which harm is caused collectively by several singly harmless actions. This will be so when the threshold between more and less harm has been exceeded. For instance, suppose that the reservoir is a triggering case, and that t_x, for some $x \in \{1, \ldots, n\text{-}1\}$, performs the triggering action. Suppose further that the number of tippers who tip their waste into the reservoir is greater than x. Given this, holding the actions of others fixed, if any given one of the tippers were to abstain from dumping their waste, the harm suffered by the villagers would be precisely the same: no threshold would be crossed. As such, each singular act of waste

disposal makes no difference to that harm. Nevertheless, it is the collective result of those actions taken together.

This demonstrates that the harmful/harmless series premise must be distinguished from a closely related claim:

The harmful/harmless plurality claim: It is possible for a plurality of actions to be both collectively harmful and individually harmless.

This claim is entailed by the harmful/harmless series premise, since a series is a plurality. However, the entailment does not go the other way; a plurality of collectively harmful/individually harmless actions is not necessarily a collectively harmful/individually harmless series of actions. Triggering cases in which the harm threshold has been exceeded provide examples of such pluralities.

Let's move now to clause (b), which claims that it is possible for there to be a collectively harmful/individually harmless series of actions such that if the relevant individuals have a reason to refrain from their collectively harmful actions, then this must be an outcome-reason. Again, the reservoir is supposed to be an example of such a case.

As Kagan points out in the passage I quoted at the outset of this section, what seems to worry us in cases like this is not the type of action which stands to be performed, but rather the outcome which those actions will jointly bring about. This is why such cases are supposed to be the favoured terrain of consequentialists: here, we are supposed to be concerned with "consequences and nothing but consequences" (Kagan 2011, p. 107). To be concrete, there doesn't appear to be anything worrisome in itself about the act of disposing of a harmlessly small amount of waste in a reservoir. What motivates the thought that these actions ought not to be performed is rather the fact that if they were, then they would cause the villagers to be poisoned. Given this, it seems unlikely that the tippers will have an action-reason to refrain from their acts of waste disposal. Rather, if there is any such consideration, it seems

most likely that it will be of the outcome variety: since it is the prospective joint outcome of these actions which concerns us.

Nor will it help matters to appeal to risk-based reasons. For any member of a harmful/harmless series of actions, the probability that that action will cause harm is zero. For instance, for any member of the series of waste-tippings, there is no chance that that action will cause harm to the villagers: since the toxicity of the reservoir's waters will be precisely the same whether they opt to tip or not. As such, the tippers cannot have a risk-based reason to refrain.

Risk-based and action-reasons do not exhaust the kinds of reason for action which are grounded in something other than the value of (actual) outcomes. The considerations raised here are thus not conclusive in favour of clause (b) of the harmless/harmful series premise. However, the principal motivation for this claim, I propose, is the thought that what primarily worries us in cases like the reservoir is the joint outcome which stands to be produced. It lies with the denier of this claim to demonstrate otherwise.

2.4. Rejecting consequentialism

Before turning to consider some possible solutions to the paradox I have set out, I want to return to a point made earlier on in this section: namely, that this paradox poses a problem for non-consequentialists. We are now in a better position to see why this might be so. The paradox of collective harm comprises of three inconsistent claims. As such, it will pose a problem for anyone who cannot easily reject at least one of those claims. The question to be addressed now it this: If consequentialism is false, will it follow that one (or more) of these claims is also false?

The answer, I think, is 'no'. The falsity of (act-)consequentialism would imply that the permissibility of an action is not determined entirely by its consequences. This proposition is clearly consistent with the truth of clause (a) of the harmful/harmless series premise; whether or not the permissibility of an action is determined by its consequences does not affect whether or not a series of actions could be both collectively harmful and severally harmless.

That the falsity of consequentialism does not undermine the truth of the other components of the paradox will be more readily appreciated, I suggest, if we consider that non-consequentialist ethical theories can (must?) allow that an action's consequences may *partially* determine its permissibility. In other words, one does not need to think that consequences should play no role in determining what people ought to do, in order to think that consequentialism is false.[89]

Given this, it becomes clear that the rejection of consequentialism need not take away from the independent appeal of the three claims which comprise my paradox. Take the collective harm premise. The case for this claim, I have argued, is abductive: harm caused collectively by many actions can, we think, be morally objectionable; and the best explanation of this, it seems, implies the collective harm premise. Nothing in this relies on the truth of consequentialism; we can think that individuals have a claim against being harmed as the joint or singular result of others' actions, even if we deny that the rightness/wrongfulness of an action is fixed entirely by its consequences.

What about the inefficacy premise? Does the falsity of consequentialism give us any less reason to adopt this claim? Again, I think not. The case I presented for this premise ran as follows: if we assume a broad conception of harm, then it is difficult to see how someone could have a reason to refrain from a harmless action which is grounded in the badness of that action's outcomes. Here again, there seems to be nothing which a non-consequentialist couldn't accept. The claim is not that we have *no* reason to avoid harmless acts. As such, it is consistent with the idea that things other than an action's consequences can be relevant to its permissibility.

[89] This point has been endorsed by some of the most prominent non-consequentialist thinkers of recent years. Rawls, for instance, writes: "All ethical doctrines worth our attention take consequences into account in judging rightness. One which did not would simply be irrational, crazy" (Rawls 1999, p. 26). David Wiggins makes a similar point about utilitarianism: "Every sane moral philosopher since the beginning of the subject has had regard for outcomes and consequences … The thing that is new about Utilitarianism is not to have dwelt on consequences – these are really nobody's monopoly – but to have sought to fix the extension of the predication 'acts rightly' *purely in terms of consequences*" (Wiggins 2006, p. 149, original emphasis).

Finally, consider clause (b) of the harmful/harmless series premise: there could be a series of collectively harmful/individually harmless actions such that if the relevant individuals have a reason to refrain from these actions, then it must be an outcome-reason. What motivates this claim, I have suggested, is the thought that in cases like the reservoir, what primarily concerns us is the outcome which the series of actions stand jointly to bring about. Again, non-consequentialists can happily accept this. Whilst a non-consequentialist will think that things other than consequences are relevant for the moral evaluation of an action, she need not deny that, at least sometimes, we should be concerned about the outcomes of actions.

3. Some solutions

We have a problem. Three claims each appear to be independently plausible. And yet they are inconsistent. These are:

The paradox of collective harm

(i) ***The collective harm premise:*** If an action is one of several which will collectively cause harm to others, then one must have a reason not to perform it.

(ii) ***The inefficacy premise:*** One has no outcome-reason to refrain from an action if, considered by itself, it is harmless.

(iii) ***The harmful/harmless series premise:*** It is possible:

(a) for a series of actions to be collectively harmful, whilst each member of that series is individually harmless; and

(b) for there to be such a series such that if the relevant individuals have a reason to refrain from their collectively harmful actions, then it must be an outcome-reason.

So, at least one of this triad must be false. But which one(s)? And why? The following subsections assess some of the answers which other theorists have put forward. The eventual conclusion will be that none of these proposals is tenable. As such, I think we have good reason to try a fresh approach, an approach which I will go on to develop in Chapter 5. I will not consider every solution which has been proposed; the following can be seen as a sort of

'Greatest Hits'. This is not, however, to say that the proposals I omit are not worth engaging with.[90] My point is simply to suggest that some of the most innovative existing solutions face serious problems; and that, given this, perhaps we should try out something quite starkly different.

3.1. Risk and vagueness

A number of authors have argued that the offending claim is the harmful/harmless series premise: or, more precisely, clause (a) of this premise. These authors claim that there could not be a series of actions which are both collectively harmful and individually harmless; that is, to employ the terminology introduced in Section 2.3, they claim that non-triggering cases are impossible.[91]

The argument in favour of this solution runs as follows. Suppose that the tippers in the reservoir example each possess one unit of waste. If we assume that this example is a non-triggering case, then we can derive the following claim: there is no number x such that if the reservoir were to contain x units of waste then its water would be safe to drink, whilst if it were to contain x+1 units then its water would be harmful to drink. In other words, there is no precise point at which one more unit of waste could make the difference between safe and harmful drinking water. It was stipulated that the reservoir's water will be safe to drink if none of the tippers tip their waste into it. Given the claim that was just derived, it follows that if only one tipper opts to tip, then the water will also be safe to drink. The problem is that we can continue this line of reasoning for two, three, four, …, tippers, until we reach the result that the water would be safe to drink if all of the tippers chose to tip. However, it was stipulated that the reservoir's water would be poisonous if all of the tippers chose to tip. Thus, our assumptions have led us into a contradiction: if all of the tippers tip, then the water will be both safe and unsafe to drink.

[90] Some notable omissions include (Barnett 2018; Hedden 2020; Otsuka 1991; Pinkert 2015; Spiekermann 2014).
[91] For views of this sort, see e.g. (Arnzentius and McCarthy 1997; Kagan 2011; Sangiovanni 2018).

This is a version of the infamous sorites paradox. As such, call this the *sorites-style argument*.[92]

The above argument is typically presented as a *reductio* of the claim that non-triggering cases are possible, and thus of clause (a) of the harmful/harmless series premise. According to advocates of this argument, all putative non-triggering cases, such as that of the reservoir, are in fact triggering cases; in such cases, there must be at least one action in the collectively harmful series which crosses a precise threshold from less harm to more.

This result would resolve the paradox of collective harm. However, a closely related problem looms. I said earlier that clause (a) the harmful/harmless series premise needed to be distinguished from what I called the harmful/harmless plurality claim: the claim that it is possible for a plurality of actions to be both collectively harmful and severally harmless. This claim is consistent with the impossibility of non-triggering cases, since a triggering case in which the harm threshold has been exceeded presents an example of a collectively harmful/individually harmless plurality of actions. However, it

[92] Nefsky (2011, pp. 384-385) considers, but ultimately rejects, an argument of this form. A slightly different type of argument is in fact more common amongst advocates of the view under consideration. Such arguments contain an additional premise which states that if two outcomes are equally harmful, then the individuals who are harmed will report on them in precisely the same way. For instance, applied to the reservoir case, such a premise would state that if there is no difference in the harmfulness of water samples containing incrementally different levels of waste, then any given villager will give the same answer to the question 'Have you been harmed?' after they have drunk each sample. The argument then proceeds in terms of such reports: we start off with 'Any given villager will report that they have not been harmed after they have drunk a sample containing 0 units of waste' and ending with 'Any given tipper will report that they have not been harmed after they have drunk a sample containing n units of waste (where n = the total number of tippers)'. This last step, however, is meant to yield a contradiction with our initial supposition that water containing n units of waste would be harmful to drink. For versions of this type of argument, see (Arnzentius and McCarthy 1997; Binmore and Voorhoeve 2006; Kagan 2011, p. 132). The plausibility of such arguments rests on the assumption that the relevant harm reports are accurate: i.e. that changes in the content of such reports correspond to changes in the level of harm inflicted on the reporters. However, if, as seems plausible, a person can be harmed without being aware of it, then this supposition will be doubtful. Perhaps it would be more plausible to assume that such reports are accurate if they were reports of pain, rather than harm (advocates of report-based arguments do in fact concentrate on cases of collectively caused pain). However, even if this assumption seems suspect (see Hedden 2020, pp. 545-546; Nefsky 2011, pp. 380-384). In order to avoid introducing questionable assumptions about their accuracy, I concentrate on a version of the argument which omits reference to harm reports.

seems we can simply substitute it into the paradox of collective harm to generate the following inconsistent set of claims:

(i) ***The collective harm premise:*** If an action is one of several which will collectively cause harm to others, then one must have a reason not to perform it.
(ii) ***The inefficacy premise:*** One has no outcome-reason to refrain from an action if, considered by itself, it is harmless.
(iii´) It is possible:
(a) for a plurality of actions to be both collectively harmful and severally harmless; and
(b) for there to be such a plurality such that if the relevant individuals have a reason to refrain from their collectively harmful actions, then it must be an outcome-reason.

If each of these claims has independent plausibility, then the sorites-style argument will have avoided my original paradox only to land us with a new one. Can this result be avoided?

A possible solution is to appeal to considerations of risk. Suppose we accept the sorites-style argument and hold that all putative non-triggering cases are in fact triggering cases. Given this, in many such cases, it seems the relevant individuals will be ignorant of whether or not their action will trigger harm to be caused. In the reservoir case, for instance, it seems that none of the tippers will know where the harm thresholds lie: that is, precisely how many waste-tippings will need to be performed in order for harm to be triggered. Such uncertainty poses a problem for rational decision making; in order to assess their options, individuals need to be able to make assumptions about what the consequences of their actions will be. One way in which to resolve this difficulty is to assume that all of the relevant collectively harmful actions are equally likely to be a triggering action. Given this, each individual can assume that their action has some small probability of bringing about a harmful outcome.

In Section 2.2, I assumed that an individual has a reason to avoid actions which pose an unacceptable risk of harm to others, a reason the strength of which is a function of the seriousness of that possible harm and the probability that the action in question will bring it about. If we grant that uncertainty about harm thresholds in triggering cases is to be resolved in the way described in the preceding paragraph, then the individuals in such cases will all have a risk-based reason to refrain from their collectively harmful actions. Depending on how many actions stand to be performed in the case at hand, the probability of a given action's being a trigger will be relatively small. However, if the harm which stands to be caused is sufficiently serious, then this reason may nonetheless have considerable strength.

I also supposed in Section 2.2 that such risk-based reasons are distinct from outcome-reasons: they are grounded not in the (dis)value of an action's actual consequences, but rather in that of its possible consequences. This opens up space for an advocate of the sorites-style argument to deny claim (iii′(b)) of the foregoing triad. If the reservoir is supposed to provide an example of the kind of case which would verify that claim, then, given the argument of the last few paragraphs, it fails; the tippers each have a risk-based reason, not an outcome-reason, to refrain from their waste-tippings.[93]

Let's take stock. The sorites-style argument purports to resolve the paradox of collective harm by rejecting clause (a) of the harmful/harmless series premise. I have raised a possible objection which alleges that this solution rescues us from one paradox only to land us with another: namely, that composed of claims (i), (ii), and (iii′). And I have sketched a possible solution to that second paradox which appeals to considerations of risk. So far, then, the prospects of the present solution look good. However, we have not yet considered whether the sorites-style argument itself is sound. Let's turn now to that argument.

[93] Kagan (2011) presents a solution along these lines. Assuming the classical utilitarian procedure for decision under uncertainty (viz. an appeal to *expected*, as opposed to *actual*, utility), he argues that triggering cases pose no problem for consequentialism. For critical responses, see (Budolfson 2019; Nefsky 2011). For a response to Budolfson, see (Hedden 2020).

The sorites-style argument, remember, purports to be a *reductio* of the claim that non-triggering cases are possible. If we assume this claim, it says, then we can derive a contradiction. For instance, if we assume that the reservoir is a non-triggering case, then we can derive the conclusion that if all of the tippers were to dump their waste in the reservoir, then its waters would be both safe and unsafe to drink. Since this conclusion is necessarily false, the assumption must be false as well.

Nefsky (2011, p. 384-386) objects that this argument relies on a particular kind of solution to sorites paradoxes in general: specifically, solutions which posit sharp extensional boundaries for vague predicates, such as epistemicism (more on this shortly). But since the jury is still out as to the best way of dealing with the sorites, we should remain neutral between the going approaches, and so should not endorse the sorites-style solution to the paradox of collective harm. This objection, I think, is on the right track. However, as it stands, it is flawed. My aim in the remainder of this subsection will be to identify where, exactly, the sorites-style argument goes wrong. Its principal error, I suggest, lies not in the assumptions it makes about the logic of vague terms, but rather in its assumptions about the structure of non-triggering cases.

It will be helpful to render the sorites-style argument formally using the following notation. Let L be a first-order language containing: the connectives '$\neg$' and '$\wedge$'; the names 'w_0', 'w_1', …, 'w_n', where w_0 stands for 'the reservoir's water when it contains 0 units of waste', w_1 stands for 'the reservoir's water when it contains 1 unit of waste', and so on up w_n, where n = the total number of tippers; the predicate H, where Ha stands for 'a is harmful to drink'; and the quantifier $\exists x$ which ranges over the integers 0, …, n.

The sorites-style argument assumes, as premises, that water containing no waste would be safe to drink, and that water containing *n* units of waste would be harmful to drink. We can render these premises in *L* as follows:

(1) $\neg Hw_o$

(2) Hw_n

The argument then supposes that if the reservoir is a non-triggering case, then there will be no number *x* such that water containing *x* units of waste would be safe to drink, whilst water containing *x*+1 units would be harmful. Given this, we are asked to assume the consequent for *reductio*. It will be helpful to have a name for this assumption; call it (NO SHARP BOUNDARIES). It can be rendered in *L* as follows:

(3) (NO SHARP BOUNDARIES) $\neg\exists x\ (\neg Hw_x \wedge Hw_{x+1})$

From (1) and (3) we can deduce:

(4) $\neg Hw_1$

From (3) and (4) we can deduce:

(5) $\neg Hw_2$

…

From (3) and (*n*+2) we can deduce:

(*n*+3) $\neg Hw_n$

But (*n*+3) contradicts (2). (1), (2), and (3) are the argument's only premises and undischarged assumptions. Since assuming (3) has led us into a contradiction, we can discharge it and infer:

(*n*+4) $\neg\neg\exists x\,(\neg Hw_x \wedge Hw_{x+1})$

Therefore, by Double Negation Elimination, we can deduce:

(*n*+5) $\exists x\,(\neg Hw_x \wedge Hw_{x+1})$

As it is the negand of (NO SHARP BOUNDARIES), let's call (*n*+5) (SHARP BOUNDARIES). The sorites-style argument now concludes as follows. Since (SHARP BOUNDARIES) follows from true premises, it must be true. But since the assumption that the reservoir is a non-triggering case implies (NO SHARP BOUNDARIES), this result is inconsistent with our assumption. So, contrary to appearances, the reservoir must be a triggering case. *QED*.

For the moment, let's put to one side the claim that (NO SHARP BOUNDARIES) must be true if the reservoir is a non-triggering case, and focus instead on the argument's explicit steps which I have rendered in *L*. Notice that (SHARP BOUNDARIES) would strike many people as surprising. Most would, at the very least, hesitate before affirming that the series of incremental changes in the reservoir's waste-level contains some sharp boundary between safe and harmful drinking water; and if anyone were asked to locate such a boundary, they would surely be unable to do so. And yet (SHARP BOUNDARIES) is derived from true premises by applications of classical logic. What, then, are we to do?

One approach is to retain classical logic and to offer an explanation as to why, whilst it is in fact true, (SHARP BOUNDARIES) *seems* to be false and (NO SHARP BOUNDARIES) true. One such account is offered by epistemicism. Epistemicists claim that whilst all vague predicates, like 'is harmful to drink', have sharp extensional boundaries, these boundaries are unknowable; we cannot know, for instance, at what point one more unit of waste would make the difference between safe and harmful drinking water. Because of such

ignorance, competent language users are apt to believe (incorrectly) that claims like (NO SHARP BOUNDARIES) are true and (SHARP BOUNDARIES) false.[94]

Notice that if epistemicism is correct, then the sorites-style solution to the paradox of collective harm would seem to be on firm ground. Epistemicists will endorse (SHARP BOUNDARIES). So, if (SHARP BOUNDARIES) is incompatible with the claim that the reservoir is a non-triggering case, then they will reject clause (a) of the harmful/harmless: all putative non-triggering cases will in fact contain at least one precise harm threshold. Furthermore, since the locations of these thresholds will be unknowable, individuals in such cases will have risk-based reasons to refrain from their collectively harmful actions.

There are, however, other ways of dealing with the sorites. According to some theorists, the best way in which to resolve the paradox is not try and reconcile the results of classical logic with the behaviour of ordinary language users, but rather to depart from classical logic. For instance, advocates of intuitionistic logics for vague terms reject unrestricted applications of the classical law of Double Negation Elimination, and so deny the move from (n+4) to (SHARP BOUNDARIES).[95]

This might seem to give us all we need to dismiss the sorties-style solution: its argument is sound if we adopt epistemicism, but not if we endorse a non-classical logic for vague terms. Yet we should not be too hasty. Remember that the sorites-style argument asserts that it is a condition of the reservoir being a non-triggering case that (NO SHARP BOUNDARIES) be true; if this claim is anything other than true, then it will not be true to say that the reservoir is a non-triggering case. However, whilst epistemicists diverge from advocates of certain non-classical logics in affirming the truth of (SHARP BOUNDARIES),

[94] For a much-discussed defence epistemicism, see (Williamson 1994).
[95] For a recent defence of an intuitionistic logic for vague terms, see (Bobzien and Rumfitt 2020).

they are not alone in holding that (NO SHARP BOUNDARIES) is not true.[96] For example, intuitionists hold that we should neither affirm nor deny (SHARP BOUNDARIES). Since this claim is not a clear truth, we should not affirm it. But denying it gives us (NO SHARP BOUNDARIES), which is incompatible with the clearly true claims (1) and (2). Since it rejects Double Negation Elimination, this ambivalence can be expressed in an intuitionistic logic by affirming (*n*+4): i.e. the negation of (NO SHARP BOUNDARIES).

What this shows is that an advocate of the sorites-style argument doesn't need (SHARP BOUNDARIES) to be true. All they need is for (NO SHARP BOUNDARIES) not to be true. This might sound strange, as the latter is the negation of the former. But remember that we are dealing here with non-classical logics; and, as I just demonstrated with the example of intuitionistic logics, in some of these, a negation can be not-true (or even false) without its negand being true.[97] But if so, then the charge that the sorites-style argument relies on one theory of vagueness in particular seems unfounded. Many views, not just epistemicism, hold that claims like (NO SHARP BOUNDARIES) are false, or at least not-true. So, if this claim needs to be true in order for the reservoir to be a non-triggering case, then a solution to the paradox of collective harm which declines to affirm clause (a) of the harmful/harmless series premise will be consistent with various ways of theorising vagueness.

Does this mean that the sorites-style solution is successful after all? Not necessarily. There is, I suggest, another way to resist the argument. Rather than trying to affirm (NO SHARP BOUNDARIES), a venture that will lead us into conflict with numerous theories of vagueness, I propose that we deny that this claim must be true in order for the reservoir to be a non-triggering case. To this end, suppose that (NO SHARP BOUNDARIES) is false; and, to make my task even harder, let's assume bivalence and suppose that (SHARP BOUNDARIES) is

[96] Note, however, that some non-classical logics for vague terms agree with epistemicism in affirming claims like (SHARP BOUNDARIES). For instance, supervaluationism implies that, whilst it has no true instances, (SHARP BOUNDARIES) is true. See e.g. (Fine 1975).
[97] According to views which reject bivalence, a sentence can be not-true without being false. These views vary in how they characterise such sentences. Some posit 'truth-value gaps' (i.e. sentences which are neither true nor false). Others posit degrees of truth and falsity.

therefore true. In other words, we are assuming that there exists some precise point at which one more unit of waste would make the difference between safe and harmful drinking water. Even so, I suggest that the reservoir could still be a non-triggering case.

Recall how I characterised non-triggering cases in Section 2.3. A case is non-triggering, I said, if it is a case the possibility of which would make clause (a) of the harmful/harmless series premise true: that is, if it is a case in which a series of actions collectively cause harm, whilst no single one of them causes any harm. Given this, in order for the reservoir to be a non-triggering case, the following needs to be true: for any tipper $t_x \in \{t_1, \ldots, t_n\}$, if t_x were to tip, then the harmfulness of the reservoir's water would be precisely the same as it was when t_{x-1} tipped (interpreting 't_0' as 'no one'). Call this claim (NO HARM THRESHOLDS). What I want to suggest is that, given certain background assumptions, (NO HARM THRESHOLDS) could be consistent with (SHARP BOUNDARIES).

One might find this suggestion surprising. How could each tipper's action be singly harmless if there exists some sharp boundary at which the addition of a single unit of waste would make the difference between harmless and harmful drinking water? Surely there will come a point in the series of waste-tippings at which one more waste-tipping would take us across that boundary? This line of reasoning would be correct if it was supposed that the boundary's location is fixed: that is, that the point at which one more unit of waste would take us from harmless to harmful drinking water is always the same. However, that assumption is contestable. Delia Graff Fara (2000), for example, has argued that, whilst vague predicates draw sharp extensional boundaries, the locations of these boundaries can shift from context to context. More specifically, she argues that the very act of considering the extreme similarity of two adjacent members of a sorites series brings it about that they both fall within the extension, or anti-extension, of the predicate in question, so that "any attempt to bring [the extensional boundary] into focus causes it to shift somewhere else" (Graff 2000, pp. 75-76).

If a view like Graff Fara's is correct, then it could be the case that a single waste-tipping never makes a difference to the harmfulness of the reservoir's water, even if there exists a sharp boundary between safe and harmful drinking water. The very act of considering whether a given waste-tipping is more harmful than its immediate predecessor in the series might cause that boundary to shift so that it does not fall between the two. Otherwise put, given the assumption that vague predicates are context-sensitive in the way described by Graff Fara, (NO HARM THRESHOLDS) will be consistent with (SHARP BOUNDARIES). But if so, then the reservoir could be a non-triggering case, even if it is supposed that (SHARP BOUNDARIES) is true.[98]

If all of this is correct, then it opens up the following response to the sorites-style solution to the paradox of collective harm. This solution does turn out to rely on substantive assumptions within the theory of vagueness (this is why I said earlier that Nefsky's objection is on the right track). However, this is not because it assumes that (NO HARM THRESHOLDS) is not true; many accounts of vagueness would agree on that point. Rather, it is because it assumes that the reservoir could not be a non-triggering case unless (NO HARM THRESHOLDS) was true. That assumption, I have argued, will be correct only if we further suppose that the extensional boundaries of vague predicates do not shift from context to context; and that is a contestable supposition. So, if a solution to the paradox of collective harm should avoid making disputed

[98] Spiekermann (2014, p. 80) also argues that one way in which to defend the possibility of non-triggering cases would be to adopt a contextualist semantics for vague terms. His point, however, differs from mine. Spiekermann's broader strategy is to block the sorites-style argument by rejecting the transitivity of 'is as harmful as'. This is an odd-sounding claim to make. Yet Spiekermann points out that it has some plausibility if we adopt a (largely discredited) hedonistic account of harm, on which harm is equated with unpleasant experiential states. Given the hedonistic account, 'is as harmful as' will be an instance of the more general relation 'is indiscriminable from', a relation whose intransitivity some authors have defended; for an ultimately critical discussion of these views, see (Graff 2001). However, even if Spiekermann is right on this point, the intransitivity of 'is as harmful as' will be much harder to accept if we drop the hedonistic account and allow that non-experiential states can be harmful. For an argument that all relations of the form 'is exactly as *F* as' are transitive, see (Hedden 2020).

assumptions about the semantics of vague terms, then we should resist the sorites-style solution and accept the possibility of non-triggering cases.[99]

3.2. Participation

How else might we try to resolve the paradox of collective harm? One prominent alternative appeals to the notion of *participation*. This approach can be sketched along the following lines:

'Granted, when an individual performs one of a series of collectively harmful/individually harmless actions, they don't cause any harm. Nevertheless, they *participate* in the collective production of harm. Moreover, we have a reason not to participate in the collective production of harm. Therefore, it is not the case that one has no reason to refrain from one of a series of collectively harmful/individually harmless actions.'

Unlike that explored in the preceding subsection, the participation approach endorses clause (a) of the harmful/harmless series premise. The component of the paradox which it proposes we reject instead will depend upon the sort of reason it claims we have to avoid participating in the joint production of harm. If it is an outcome-reason, then the solution will reject the inefficacy premise; we can have an outcome-reason to refrain from a harmless action. If, on the other hand, it is another sort of reason, then the solution will deny clause (b) of the harmful/harmless series premise; we can have a reason to refrain from actions like the tippers' acts of waste disposal, even if it is not an outcome-reason. For present purposes, it does not matter which of these options we pick, and so I will leave both open.

Nefsky (2015) distinguishes between two possible versions of the participation approach: *weak participation* and, *strong participation*. Views of the former type offer a minimal interpretation of what it is to participate in

[99] Zach Barnett (2018), Brian Hedden (2020), and Michael Otsuka (1992) have each raised criticisms of claims similar to clause (a) of the harmful/harmless series premise which do not rely on the sorites-style argument. Interesting as they are, I will leave it for another time to discuss them.

the collective production of harm, whilst the latter offer a more substantive interpretation. The following subsections will elaborate and evaluate each of these approaches respectively.

3.2.1. Weak participation

According to weak participation views, an individual participates in the joint production of a harm if and only if their action belongs to a particular set of actions. Variants of this approach then differ according to how they identify this set.

There are two minimal conditions which this identification criterion must satisfy, in order to yield a credible solution to the paradox of collective harm. First, it must correctly describe the members of the collectively harmful/individually harmless series of actions which stand to be performed in non-triggering cases. If the criterion does not meet this condition, then the resulting weak participation view will be unable to deliver its promised solution to the paradox of collective harm: since it will not deliver the result that we have a reason to refrain from such actions. Second, the identification criterion must identify a quality of the relevant actions which plausibly gives us a reason not to perform them. This second condition is important: since if it is not satisfied, the view's assumption that we have a reason not to participate in the joint production of harm will be question begging.

How, then, are we to identify the set of actions, the members of which participate in the joint production of harm? Nefsky interprets the following claim from Derek Parfit as an answer to this question: "Even if an act harms no one, this act may be wrong because it is one of a set of acts that together harm other people" (Parfit 1984, p. 70). According to Nefsky, we can read this claim as follows. We have a reason not to participate in the joint production of harm (indeed, it is morally wrong to do so); and an individual participates in the joint production of harm if their action belongs to a set of actions the members of which together cause harm to others.

This proposal satisfies the first condition: by stipulation, actions like the tippers' waste-tippings collectively cause harm to others. But does it satisfy the second? The conjunction of the inefficacy premise and the harmful/harmless series premise seems to imply that it does not. The former premise says that we have no outcome-reason to refrain from individually harmless actions. But the latter implies: (a) that several actions can together cause harm, even if each is individually harmless; and (b) that the relevant individuals could lack any reason to refrain from such actions which is not of the outcome variety. It follows that the actions of several individuals can jointly harm others, without any of them having a reason not to perform them.

What this demonstrates is that, in order for Parfit's proposal to deliver a solution to the paradox, either the inefficacy premise or the harmful/harmless series premise must be false. But we have yet to see an argument for any of these claims. Thus, in the absence of such further argument, Parfit's weak participation view begs the question; it assumes, in the face of an argument to the contrary, that we have a reason not to perform actions which together cause harm to others.

I will return to the weak participation approach later on. For now, however, let's turn to the strong participation approach.

3.2.2. Strong participation and superfluity

In contrast to weak participation views, strong versions of the participation approach offer a substantive account of what it is to participate in the collective production of harm. Nefsky finds an example of such a view in Christopher Kutz's (2000) work on complicity. According to Kutz, an individual participates in the joint production of some harmful outcome *h* if she is one of the agents of a joint action *j*, and *h* is an is outcome of *j*. As such, the notion of participation is equated with that of joint agency.

Kutz develops an account of joint agency on which an individual counts among the agents of a joint action only if she acts with a "participatory intention" (Kutz 2000, p. 74): that is, only if she intends her own singular

action as a contribution to some collective end, where a 'collective end' is either "constituted by or … a causal product of different individuals' acts" (Kutz 2000, p. 81-82).[100] Thus, in order for any given one of the tippers to participate in the joint poisoning of the villagers, they must act towards the accomplishment two goals: first, they must intend to tip their own waste into the reservoir; and second, they must simultaneously intend that singular action as a contribution to the collective end of poisoning the villagers.

An immediate problem with Kutz's approach is that cases like that of the reservoir can be re-described such that no one would act with the intention of contributing to the joint end of poisoning the villagers. For instance, we could suppose that each tipper is offered £10 to dispose of their waste in the reservoir; and that if any of them chose to do so, they would act only with the intention of enriching themselves. In this case, Kutz's account seems to imply that even if all of the tippers were to dump their waste in the reservoir, none of them would have participated in the collective production of harm: since, whilst each may have intended to tip their waste, none will have intended their singular act as a contribution to a collective end of harming the villagers. As such, in this modified case, the account will not imply that tippers would act unreasonably by tipping their waste into the reservoir. And yet it seems the harm caused could nonetheless be morally objectionable. As such, the abductive grounds for thinking that the tippers do in fact have a reason refrain remain.

Nefsky (2015, pp. 251-252), however, offers a response. She suggests the following amendment to the strong participation view: if an individual could feasibly participate in a joint effort to prevent harm from occurring, then they have a reason to do so. Presumably, each of the tippers could feasibly refuse the £10; and, moreover, they could do so with the intention of contributing to the collective end of preventing the villagers from being harmed. As such, coupled with Kutz's account of participation, Nefsky's amendment seems to

[100] For Kutz, acting with a participatory intention is necessary but not sufficient for participating in a joint action. In addition, there must be sufficient overlap between the relevant individuals' conceptions of the joint action (Kutz 2000, pp. 94-95).

imply that the tippers each have a reason not to dump their waste in the reservoir, even in the re-described case.

Nevertheless, Nefsky argues that Kutz's approach faces a problem, which she labels the *superfluity problem* (Nefsky 2015, p. 259). At base, the problem is this: if an action, considered singly, makes no difference to a particular outcome, then it is natural to think that it is superfluous with respect to it. The notion of superfluity is difficult to pin down, and it will be explored more deeply in Section 3.4.1. Roughly speaking, however, actions which are superfluous with respect to a given outcome are ones which do not help to bring that outcome about. To see the naturalness of the preceding inference, consider the following example.[101] Suppose that an undergraduate in her final year calculates that, given the grades she has already received, she will receive a First regardless of what she scores on her final exam. Whether or not she scores highly on this exam makes no difference to whether or not she will receive a First overall. Given this, it seems valid for the student to infer that attempts to improve her score would be superfluous with respect to her receiving a First. That is, acting in this way will not help to bring this outcome about; and this seems to be so precisely because it will make no difference to whether or not it occurs.

Nefsky (2015, pp. 261-263) argues that this inference, if valid, creates the following problem for Kutz. Suppose that several agents share a collective end to bring about some outcome *o*. Nefsky claims that an individual action could coherently be described as a contribution to that collective end only if it is not superfluous with respect to *o*. In other words, if it is our joint goal to bring about some state of affairs, and my φ-ing will not help to bring that state of affairs about, then if I φ, I cannot rationally intend my act as a contribution to our collective end. For instance, suppose that you and I share the joint goal of moving a piano down a flight of stairs. Presumably, my twiddling my thumbs is superfluous with respect to that end; this act will not help to shift the piano. As such, it seems that I cannot rationally intend my twiddling as a

[101] The example is adapted from (Nefsky 2017, pp. 2758-2759).

contribution to our joint end; if I were to say that I participated in the piano moving by twiddling my thumbs, then you would rightly say that I am mistaken.

But if this is correct, then the superfluity problem will undermine the solution to the paradox of collective harm which Kutz's notion of participation is supposed to provide. If the fact that an action makes no difference to some harmful outcome implies that it is superfluous with respect to it, then it will follow that individuals in non-triggering cases will be unable rationally to intend their collectively harmful/individually harmless actions as contributions to a joint end of inflicting harm. Nor will they be able rationally to intend their abstentions as contributions to a collective goal of preventing harm to others. As such, the tippers, for instance, will have no reason grounded in the notion of participation to refrain from dumping their waste into the reservoir.

Again, Kutz's strong participation approach will be revisited at a later point in this chapter. But in the meantime, let's consider a further possible solution.

3.3. Fairness

A different sort of approach stems from the thought that it would be unfair for an individual to perform one of a series of individually harmless, though collectively harmful, actions.[102] This thought, if correct, suggests an alternative solution to the paradox of collective harm. We seem to have an action-reason to refrain from doing unfair things; that is, we appear to have a reason not to act unfairly *as such*, regardless of the further consequences which tokens of that act-type might have. Thus, if it is unfair to perform one of a series of individually harmless/collectively harmful actions, then clause

[102] Adam Swift (2003, pp. 147-148), for instance, appeals to this thought in order to argue that parents should send their children to (adequately good) local comprehensive schools, rather than to elite private schools. If many middle-class parents were to move their children to private schools, he claims, then this would have a detrimental effect on the children who remain in the comprehensives (because of the beneficial influence which both the parents and their children can have). Given this, Swift argues that it would be unfair for a given middle-class parent to remove their child from the comprehensive system: even if, on its own, it will make no difference to the quality of education which others receive.

(b) of the harmful/harmless series premise will be false: we can have a reason to refrain from one of a series of collectively harmful/individually harmless actions which is not of the outcome variety.

3.3.1. Fairness *tout court*

Why would it be unfair to perform one of a series of collectively harmful/individually harmless actions? Why, for instance, would one of the tippers act unfairly, if they were to dump their waste into the reservoir? Typically, considerations of fairness are thought to arise when a group of individuals engage in a cooperative effort, thereby incurring certain sacrifices, in order to produce some benefit. Under such conditions, a principle of fairness will state that if an individual accepts some share of that benefit, then they should also bear some of the costs of its production. As such, a person acts unfairly when they acceptingly benefit from the cooperative efforts of others without sharing in that effort.[103]

Given this, it will be unfair for a given individual to perform one of a series of collectively harmful/individually harmless actions if and only if the following conditions hold: (a) other individuals have incurred certain costs in order to produce a benefit; and (b) by performing the action, the individual in question accepts that benefit, but does not incur any of the costs of producing it. Let's focus again on the reservoir case. In virtue of what could conditions (a) and (b) hold of this case, so that it would be unfair for a given tipper to dump their waste into the reservoir?

---

[103] This is the basic idea behind Rawls' principle of fairness. Explaining this principle, he writes:

> The main idea is that when a number of persons engage in a mutually advantageous cooperative venture according to rules, and thus restrict their liberty in ways necessary to yield advantages for all, those who have submitted to these restrictions have a right to a similar acquiescence on the part of those who have benefitted from their submission. We are not to gain from the cooperative labors of others without doing our fair share (Rawls 1999, p. 96)

For an instructive critical discussion of Rawls' principle of fairness, see (Simmons 1979, chap. 5).

One might propose the following. If the tippers abstain from dumping their waste in the reservoir, then each will incur a minor inconvenience. But if enough of them abstain, they will also collectively produce a benefit: the villagers will not be poisoned. So, under such circumstances, condition (a) is met. Furthermore, if a given tipper were then to choose to dump their waste in the reservoir, then they would accept that benefit, but without incurring any of the costs which others bore in order to produce it. This individual would thereby satisfy condition (b), and so their behaviour would be unfair.

However, this proposal faces two difficulties. First, why does preventing harm to the villagers benefit a tipper who chooses to dump their waste in the reservoir? If the level of waste present in the reservoir's waters is maintained at a non-toxic level, then this seems primarily to benefit the villagers: they are the ones who have to drink it. But it is not immediately clear how this benefits any of the tippers.

Second, even if it is internally sound, the above proposal cannot deliver the result that each of the tippers always has a reason to abstain. Understood in the way presented above, considerations of fairness provide us with reasons to behave in certain ways which are conditional on how others act; an individual's behaviour will be unfair only if others undertake to produce some benefit from which that individual stands to gain. Applied to the case of the reservoir, this implies that a given act of waste-tipping will be unfair only if sufficiently many other tippers abstain to prevent the villagers from being harmed. Thus, if all of the tippers were to dump their waste in the reservoir, causing the villagers to be poisoned, then none of them would have acted unfairly: since there would be no cooperatively produced benefit for any of them to gain from.

This second difficulty is particularly problematic, since it reveals that considerations of fairness will count against acts of waste-tipping only in circumstances where the actions of some subset of the tippers have already secured the villagers' safety. If the health of the villagers is identified as the relevant cooperatively produced benefit, then when this good is not brought

about, acts of waste-tipping will not be unfair. But it is precisely in circumstances where the tippers' actions jointly cause harm that the abductive grounds for thinking that they have a reason not to act as such arise; we find the harm they cause to be morally objectionable, and this seems to be best explained by supposing that they have a reason to abstain.

3.3.2. Fairness and collective obligation

Garrett Cullity (2000) offers a view which aims to avoid both of these two difficulties. He supplements considerations of fairness with an appeal to the notion of *collective obligation*. In the sense intended here, a collective obligation obligates not any one individual, but rather a *plurality* of individuals. Thus, if Athos, Porthos, and Aramis bear a collective obligation to do something, this does not necessarily imply that any one of them is obligated to do anything. Nor does it follow that some entity made up of Athos, Porthos, and Aramis, some 'group agent', is obligated to do something. On this conception, the subjects of a collective obligation are necessarily several in number. This way of construing the notion of collective obligation will be discussed further in Sections 3.4.3 and 3.5. For now, however, it will simply be assumed.

On Cullity's view, if, by enacting some combination of actions, several individuals can prevent some other/s from being harmed, then they are collectively obligated to do so. Thus, in the case of the reservoir, the tippers are collectively obligated not to dump their waste in the reservoir: since, if they all abstain, the villagers will not be poisoned. Furthermore, Cullity supposes that a person benefits when their obligations, both individual and collective, are offloaded. Put roughly, the thought here might be that we each have some kind of 'moral ledger', and that the balance of 'positive' and 'negative' entries in this ledger comprise an element of an individual's wellbeing; good deeds are good for the person who performs them, whilst bad deeds are bad for them. An obligation offloaded counts as a positive entry in this ledger, and so benefits the person bound by it. Given all of this, we can suppose that if some sufficiently large subset of the tippers were to secure the villagers' safety by each abstaining from dumping their waste, then all of the

tippers would benefit; the subset's actions would successfully offload the collective obligation borne by all, thereby benefiting everyone who was bound by it.[104]

This resolves the first difficulty. It is now clear how each of the tippers benefit when the poisoning of the villagers is prevented. As such, if one of the tippers chooses to dump their waste in the reservoir, while the abstinence of others has saved the villagers from being harmed, then they will be acting unfairly; in so acting, they gain from the efforts of those others, without bearing any of the productive costs.

What about the second difficulty? The problem, recall, is that acts of waste-tipping will be unfair only when the villagers' health has already been secured. If enough of the tippers dump their waste in the reservoir to cause the villagers harm, then none of these actions will be unfair: since there is no cooperatively produced benefit from which the tippers stand to gain. Cullity proposes to resolve this difficulty as follows. Whilst it would not be unfair for all of the tippers to dump their waste into the reservoir, by doing so, they would fail to offload their collective obligation. As such, when the villagers' health has not already been secured, the tippers have a collective reason to abstain: that is, a reason which counts against the waste-tippings of the plurality $t_1,\ldots, t_n$.

Cullity thus provides an attractive version of the fairness approach, one which seems to avoid the two problems which arose in the preceding subsection. However, Nefsky (2015, pp. 255-259) argues that, like the strong participation approach, Cullity's view runs into the superfluity problem. Suppose that a group of individuals undertake some cooperative effort from which I acceptingly benefit. Suppose further that if I were to try and join in

[104] I don't suppose that this picture will be entirely convincing. But, for the sake of argument, I will assume that it is plausible enough. A further question which this account raises is this: Why think that a tipper who refuses to abstain from dumping their waste into the reservoir *accepts* the benefit bestowed on them by the abstentions of the other tippers? Again, in the interests of moving the discussion on, I will pass over this issue. For a helpful discussion of what it would take for an individual to accept a cooperatively produced benefit, such that they would thereby be bound by an obligation of fairness, see (Simmons 1979, pp. 118-136).

their effort, my act would be superfluous with respect to the aim of producing the relevant benefit; whatever I do, my individual effort will not help to produce that good. Given this, Nefsky claims that it could not be unfair for me not to join in the cooperative effort; since my individual effort would be superfluous, it would not reduce the productive costs which others would have to bear. But now the superfluity problem looms large. This problem, recall, is that the fact that an action makes no difference to an outcome seems to imply that it is superfluous with respect to that outcome. If this inference is valid, then, given Nefsky's claim, the tippers' acts of waste-tipping could not be unfair. Since each one of these actions makes no difference to whether or not the villagers are poisoned, the superfluity problem implies that each is superfluous with respect to that outcome. But if this is true, then it would seem that the tippers could never act unfairly by dumping their waste into the reservoir: even if the abstentions of some have already secured the villager's safety.

3.4. Helping, without making a difference

So, the superfluity problem raises difficulties for both the strong participation and the fairness approaches. As part of her own solution to the paradox of collective harm, Nefsky (2017) attempts to defuse that problem. She argues that the fact that an action makes no difference to whether some outcome occurs does not imply that it is superfluous with respect to that outcome.

3.4.1. Defusing the superfluity problem

In order to grasp the distinction between difference-making and non-superfluity, Nefsky claims that we need a better understanding of the conditions under which an action is superfluous with respect to an outcome. Delineating these conditions, however, is a subtle matter.

To begin, it seems that an action cannot help to bring an outcome about, unless it could be part of its cause. For instance, suppose I park my car in a pay-and-display carpark. I do not want to receive a parking fine. Towards this end, I sit in my seat, twiddling my thumbs. My thumb twiddling seems superfluous with respect to my aim of not receiving a fine. Why? Because my

doing so cannot cause, nor can it be part of what causes, the parking attendant not to fine me.[105]

However, being part of what causes an outcome is not sufficient for an action to be non-superfluous with respect to it. Suppose I put some money in the parking meter and head off into town. Part way through my errands, however, I realise that I cannot remember how much parking time I paid for. I know that I'll need at least another hour, and so I return to my car to check my ticket. On checking, I see that I have already paid for an extra hour. Moreover, once this hour is up, parking becomes free. Nonetheless, I decide to pay for a further hour's parking. My paying for this extra hour could be part of what causes me not to receive a fine. If a parking attendant were to check my ticket, their reason for not fining me would be that I have paid for two more hours' parking. Yet my act seems superfluous with respect to my aim. Why is this? Nefsky (2017, p. 2752) suggests it is because when I insert my money into the meter, it is already guaranteed that I will not receive a fine. More generally, an action *a* is superfluous with respect to an outcome *o* if, when *a* is performed, *o*'s occurrence is already guaranteed

Yet, again, this condition is not necessary for superfluity. To illustrate, suppose that I have parked my car in a disabled bay, and do not possess the necessary badge. Thus, when I pay for my extra hour's parking, it is an open possibility that I will receive a fine. Nevertheless, my paying still seems superfluous with respect to my aim. Why? The answer, says Nefsky (2017, p. 2752), is that it is not possible that I will receive a fine *because* I have not paid for my stay. Rather, if I am fined, it will be because I am not entitled to park in a disabled bay.

On the back of these observations, Nefsky (2017, p. 2753) proposes the following necessary and sufficient conditions for an action to play a non-superfluous causal role in bringing some outcome about. Supposing that an

[105] Nefsky (2017, p. 2754, fn. 22) points out that an action which is not part of the cause of some outcome could nonetheless help to bring it about, if that action is constitutive of the outcome. However, for present purposes, this exception can be put to one side.

action *a* can be part of what causes an outcome *o* to occur, *a* is non-superfluous with respect to *o* if and only if, when *a* is performed, it is an open possibility that *o* will fail to occur because of a lack of actions of some type *A*, where *a* is of the type *A*. Thus, where my putting money into the meter can be part of what causes me to avoid a fine, my doing so can help to achieve my aim just when it is an open possibility that I will be fined because not enough acts of putting money in the meter have been performed.

At this point, a possible objection suggests itself. Modifying the parking example, suppose that I return to my car to find that I must pay for an extra hour's parking if I am to avoid a fine. Nefsky's conditions might seem to imply that it might help me to avoid a fine if I were to pay for only half an hour's parking: since there is an open possibility that I will be fined because of a lack of actions of that type (putting money in the meter). Yet this result seems off. Since I need to pay for a full hour's parking in order to avoid a fine, paying for only half an hour seems unhelpful.

However, in this modified case, it's not clear that my action satisfies the supposition of Nefsky's conditions. That is, it's not clear that my action could be part of what causes me to avoid a fine. If I were to pay for only half an hour, and a parking attendant were to check my ticket, they would fine me. So, here, my act could not be part of the cause of my avoiding a fine, since that state of affairs would not have occurred. Things are different if we assume that there is a possibility that I, or some other individual/s, would pay for my remaining half an hour's parking. In this case, my paying for the first half an hour could very well be part of what causes me to avoid a fine. But if we make this assumption, then it also seems that my act could help to bring that outcome about.

So, we have an account of the conditions under which an action plays a non-superfluous causal role in bringing some outcome about. If this account is correct, then the superfluity problem will be defused: the fact that an action makes no difference to an outcome will not imply that it is superfluous with respect to it. To see this, let's return to the case of reservoir. By stipulation, a

single act of waste-tipping makes no difference to whether, or to what extent, the villagers are poisoned. However, such an act can be part of what causes the villagers to be poisoned; if they are poisoned, then this will be the joint result of many waste-tippings. So, the supposition of Nefsky's conditions is satisfied.

Now suppose that a given tipper *t* is deciding whether or not to tip. Suppose further that, at the time of *t*'s decision, it is not certain whether or not the villagers will be poisoned; not enough of the others have already tipped to cause harm, and it is not guaranteed that enough will tip in the future. As such, there is an open possibility that the villagers will not be harmed. Moreover, if they are not harmed, this will because a sufficiently small number of waste-tippings have been performed. Therefore, under such conditions, *t*'s act of waste-tipping could help to poison the villagers: even though it cannot make a difference to whether that outcome occurs.

Notice finally that whilst Nefsky's account defuses the superfluity problem, it can nevertheless accommodate the thought which initially motivated the problem. More specifically, Nefsky can readily account for the fact that, typically, when an action makes no difference to an outcome, it is superfluous with respect to it. To see this, consider again Section 3.2.2's example of the student who is guaranteed a First overall, regardless of the result she receives in her final exam. Initially, we wanted to explain the superfluity of the student's studying for this exam with respect to her receiving a First by appeal to the fact that this act makes no difference to that outcome. But, given Nefsky's account, we can now see that both the *explanandum* and the putative *explanans* are in fact explained by the fact that the student's degree classification is already guaranteed. That is, the fact the student is guaranteed a First explains why revision efforts would be superfluous with respect to that outcome. But it also explains why that act would make no difference to the degree classification. More generally, we can account for the initial appeal of the inference 'if an action makes no difference to an outcome, then it is superfluous with respect to it' by saying that, in many cases, the truth of both

the antecedent and the consequent are explained by something further: namely, the fact that the outcome in question is settled in advance.

3.4.2. Nefsky's solution

Nefsky's defusal of the superfluity problem opens up a number of possible solutions to the paradox of collective harm. Since they are both hampered by that problem, its defusal allows us to revisit the strong participation and fairness approaches. This will be done in the following subsection. However, employing her account of non-superfluous causal contribution, Nefsky also proposes her own, distinct solution. That will be the focus of this subsection.

Nefsky (2017, p. 2766) presents her account as a revisionary interpretation of the 'bringing about' relation which an action bears to its outcomes. On the standard interpretation, an action brings about some state of affairs if and only if it makes a difference to it: that is, if and only if, holding other actions fixed, that state of affairs would be different in some way, depending on whether or not the action was performed. According to Nefsky, by contrast, an action can play a non-superfluous casual role in bringing about some state of affairs, even if it makes no difference to it. As such, Nefsky can be read as offering a non-standard account of the states of affairs which count among an action's consequences, one which is more inclusive than the standard 'difference-making' conception.[106]

This non-standard conception of consequences opens up a number of possible solutions to the paradox of collective harm. One option, which Nefsky does not explicitly advocate, would be to reject clause (a) of the harmful/harmless series premise: the claim that there could be a series of actions which are both collectively harmful and severally harmless. The case for rejecting this component of the paradox would not, however, be the sorites-style argument

[106] Two points. First, it is worth noting that it is not only non-triggering cases which pose a problem for the standard, difference-making conception of consequences. This conception will also run into difficulties in cases of overdetermination. For present purposes, however, I will put this further issue to one side. Second, Spiekermann (2014, pp. 86-89) has also sketched the outlines of a solution to the paradox of collective harm which also assumes a non-standard interpretation of an action's consequences. As such, I view his and Nefsky's accounts as variations on a common theme.

discussed in Section 3.1. Rather, it would be that, given Nefsky's non-standard account of consequences, the members of a collectively harmful series of actions will not turn out to be individually harmless. For instance, in the reservoir case, if the villagers are harmed as a joint result of the tippers' waste-tippings (and, when those actions were performed, it was an open possibility that the villagers would not be harmed because sufficiently few waste-tippings had occurred), then that outcome will count as a consequence of each one of those actions: each waste-tipping will play a non-superfluous causal role in the villagers' poisoning. As such, provided that Nefsky's conditions for non-superfluous contribution hold, the reservoir will not provide an example of a collectively harmful/severally harmless series of actions.[107]

A different option, one which Nefsky herself seems more explicitly drawn to, would be to reject the inefficacy premise: the claim that we have no outcome-reason to avoid actions which are individually harmless. The grounds for rejecting this premise would run as follows. We could simply stipulate that an action is individually harmless if and only if it makes no difference to any harmful outcome; holding other actions fixed, there will be no more or less harm suffered, depending on whether or not that action is performed. However, given Nefsky's non-standard conception of consequences, an individually harmless action, understood as such, could nonetheless have harmful consequences. If an action plays a non-superfluous causal role in bringing about a harmful state of affairs, then that state of affairs will count among the action's outcomes: even if that action makes no difference to it, and so is individually harmless. So, if we have an outcome-reason to refrain from actions with harmful consequences, then we could have an outcome-reason not to perform an individually harmless action.

These solutions to the paradox have many attractive qualities. They resolve the problem elegantly, whilst proposing minimal revisions to our common-

[107] If, on the other hand, those conditions do not hold, then Nefsky could claim that the case is an exception to the collective harm premise. See the final three paragraphs of Section 2.1.

sense conception of practical rationality. The only revision required is Nefsky's non-standard conception of consequences. But adopting this conception is hardly a very great price to pay; indeed, we should be willing to modify the standard view if doing so yields theoretical gains.

Despite its attractions, however, I suggest that there is a weak point in the view. Both of the possible solutions I have just outlined pivot on a common assumption: namely, that we have a reason not to perform actions which make a non-superfluous causal contribution to some harmful outcome. The first solution requires this assumption in order to ensure that the possibility of non-triggering cases is consistent with the collective harm premise; showing that the tippers' collectively harmful actions are also severally harmful will not help matters unless those individuals therefore have a reason to refrain from them. And the second solution in fact requires a finer-grained version of this assumption: unless we have an *outcome*-reason to avoid actions which make a non-superfluous causal contribution to harm, Nefsky's non-standard conception of consequences will not imply that an individual could have an outcome-reason to refrain from an individually harmless action.

So, what reason do we have to refrain from making non-superfluous contributions to harmful outcomes? Happily for the second solution, the most obvious candidate seems to be of the outcome variety: we should refrain from making such contributions to harmful outcomes because of the disvalue realised by those outcomes. However, whilst it might seem initially attractive, I suggest that this proposal lacks plausibility. As we've seen, for Nefsky, an action can help to bring an outcome about, even if it makes no difference to it. But if a harmful outcome will be just the same regardless of whether a given action is performed, then why should the badness of that outcome count against its being enacted?

To put the point another way, consider why it is that we think that the disvalue of a harmful outcome can count against the performance of actions which bring it about. The point seems so obvious as to be almost trivial: typically, by refraining from an action which would otherwise have a harmful outcome,

we can prevent that outcome from occurring; and harm, we think, is something to be prevented, if possible. However, if we adopt Nefsky's non-standard conception of consequences, then this apparently trivial point becomes entirely untrivial; indeed, it becomes false. An individual cannot be guaranteed that, by refraining from an action which would otherwise have a harmful outcome, they will prevent that outcome from occurring. If the action in question will make no difference to the harmful outcome, then it could not be prevented by refraining from that action. As such, whilst it is clear why the disvalue of a harm counts against the performance of actions which make a difference to it, it is less clear how this works with actions which merely help to bring harm about, without making a difference to it.

Perhaps the reason we have to avoid making non-superfluous contributions to harmful outcomes is grounded in something other than the disvalue of those outcomes. This proposal will be explored, in an indirect way, in the proceeding subsection. For now, however, it is enough to note that the claim that the badness of a harmful outcome counts against the performance of actions which merely help to bring it about, without making a difference to it, is less than obvious.[108]

Nefsky (2017, pp. 2766-2767) has a response to objections of this sort. If we concentrate exclusively on cases in which harm is caused by a single action, she suggests, then it might seem plausible to think that we only have outcome-reasons to refrain from actions which make a difference to harmful states of affairs. However, once we turn our attention to non-triggering cases, in which a series of actions can collectively cause harm without any one of them making a difference to it, then this thought becomes much less plausible: since it would leave us unable to say that the individuals in question have a reason to refrain from those actions, and thus unable to account for the intuitive judgement that the harm caused would be morally objectionable. However, this response relies on Nefsky's being the only workable solution

[108] James Fanciullo (2020) has also questioned Nefsky's assumption that we have a reason not to make non-superfluous causal contributions to harmful outcomes, even if those contributions make no difference. His arguments, however, differ from mine.

to the paradox of collective harm. In the following chapter, I aim to provide another.

3.4.3. Participation and fairness revisited

So, Nefsky's own solution to the paradox of collective harm faces a difficulty, even if her defusal of the superfluity problem is successful. But that defusal opens up further possible solutions. The superfluity problem posed a common challenge to both the strong participation and the fairness approaches. As such, Nefsky's conception of non-superfluous causal contribution, as distinct from difference-making, enables us to revisit those accounts.

Before doing so, however, I want briefly to reconsider the weak participation approach. Whilst I did not argue that the superfluity problem created difficulties for this view, Nefsky's account might nonetheless be used to develop it. On this view, recall, we have a reason not to participate in the joint production of harm, where one so participates just if one's action is a member of the relevant set. The challenge facing this view was that of finding an identification criterion for the relevant set which: (a) correctly describes the members of a collectively harmful/individually harmless series of actions; and (b) identifies a feature of the relevant actions which plausibly gives us a reason to avoid them.

Employing Nefsky's account of non-superfluous causal contribution, we can offer the following: an action is a member of the relevant set if and only if it plays a non-superfluous causal role in bringing about some harmful outcome. As we saw in Section 3.4.1, this criterion satisfies condition (a). However, if my claims in the preceding subsection hold, it is not clear that it satisfies condition (b). That is, it is not evident that we always have a reason to refrain from actions which play a non-superfluous causal role in bringing about harm. What we need, then, is an account which demonstrates why we should avoid such actions. Perhaps either the strong participation or the fairness view might provide this.

Consider first the strong participation approach. According to the version of this approach drawn from Kutz, one has a reason both to participate in joint efforts to prevent harm, and not to participate in joint efforts to inflict harm. Moreover, one so participates only if one acts with the intention of contributing to such a joint effort. The problem facing this view was that it seems that if an agent knows that her action is superfluous with respect to some collective goal, then she cannot rationally intend it as a contribution to that goal. But if an action which makes no difference to whether or not an outcome occurs is therefore superfluous with respect to it, then Kutz's approach will not deliver the result that we have a reason to refrain from one of a series of collectively harmful/individually harmless actions.

Nefsky's defusal of the superfluity problem allows us to block this last inference. An action which makes no difference to some outcome is not necessarily superfluous with respect to it. So, one of the tippers, for instance, could rationally intend their singular waste-tipping as a contribution to the collective end of poisoning the villagers. Furthermore, they could also rationally intend their abstaining from tipping as a contribution to the joint end of saving the villagers from being harmed. If so, then it seems that the strong participation account can deliver the conclusion that each tipper has a reason not dump their waste into the reservoir.

So far, so good; the attempt to save the strong participation approach from the superfluity problem seems a success. However, even granting this fix, I suggest that there are further difficulties facing the account. Recall the second difficulty which, in Section 3.3.1, I claimed faced the fairness approach. The issue was that considerations of fairness establish reasons which are conditional on how others behave. This creates a problem because it implies that, for instance, if all of the tippers dump their waste in the reservoir, thereby poisoning the villagers, none of them will have acted unfairly. A similar problem, I suggest, faces the strong participation approach.

To see this, suppose that all of the tippers opt to tip. On the strong participation view, the tippers will have participated in the joint poisoning of

the tippers, and thus have contravened a reason they had to refrain, only if either they intended their waste-tippings as contributions to the collective end of poisoning the villagers, or they could feasibly have abstained with the intention of thereby contributing to the collective end of saving the villagers from harm. Earlier, in Section 3.2.2, I pointed out that the first disjunct could easily be imagined to be false: for instance, if we supposed that the tippers were each paid to tip, and that they aimed solely at self-enrichment in doing so. However, in the event that all of the tippers opt to tip, it is also hard to see how the second disjunct could be true. In order for a given tipper rationally to intend their abstention as a contribution to a collective end of protecting the villagers, they must have some grounds to believe that such a collective end exists. But a collective end, recall, is either "constituted by or … a causal product of different individuals' acts" (Kutz 2000, p. 81-82). As such, if everyone else has decided to tip their waste into the reservoir, it is hard to see what grounds a given tipper could have for believing that there exists any joint end of protecting the villagers to which they might contribute.

More generally, whether or not an individual can rationally intend their singular action (or omission) as a contribution to some collective end *E* depends on whether or not others act in a way which gives that individual grounds for believing that *E* exists. As such, when others do not act in that way, that individual will not be able to participate in *E*, and so could not have participation-based reasons for contributing to it. This creates a problem where *E* is, for instance, the joint goal of protecting the villagers from being poisoned. In order for a given tipper to have grounds to believe that this joint goal exists, at least some of the other tippers must act in a way which suggests that they intend to prevent the villagers' poisoning. But it is difficult to infer any such intention when all of the tippers opt to tip. So, like the fairness (*tout court*) approach, it seems that Kuz's strong participation approach will imply that the tippers do not act unreasonably when all of them choose to dump their waste into the reservoir.

Kutz has a reply. In cases where several individuals collectively cause harm without participating in a collective end, he argues that there are two further

possible grounds for claiming that those individuals nonetheless have a reason to refrain. First, whilst there is no shared goal in such cases, it may be coherent to describe the actions in question as instances of *quasi*-participation in the joint production of harm (Kutz 2000, p. 186). As an example, Kutz refers us to Western car owners, whose actions jointly cause environmental damage. These individuals do not share a collective goal of inflicting harm on others. Yet they do, he claims, engage in a "way of life" which "treats collective and distant harms as off the moral map" (Kutz 2000, p. 186). As such, whilst none of these individuals could rationally intend their individual actions as contributions to the joint end of causing environmental damage (since there is no such end), those actions can be described as quasi-contributions to a harmful way of life. The second ground rests on the symbolic features of collectively harmful acts. Kutz writes: "agents can have reason to refrain from participating in a harm, not because of the relation between this choice and an actual outcome, but because of what the choice symbolizes in their characters and commitments" (Kutz 2000, p. 190). Even if an individual has no participation-based reason to refrain from one of a series of collectively harmful/individually harmless actions, they may nonetheless have a reason to refrain which is based on the message which performing that action might express to others.

However, as Andrea Sangiovanni (2018, pp. 471-472) has pointed out, neither of these proposals seems entirely convincing. First, even granting that we can make sense of the notion of quasi-contribution, this idea seems unhelpful when we turn to fictional cases like that of the reservoir. The example could be redescribed such that the tippers cannot be thought of as engaging in a shared way of life which treats collective harms with indifference. They could be very concerned about the harm which the villagers stand to suffer; and yet it seems that each could nonetheless doubt that they have any reason not to dump their waste into the reservoir. Second, the proposal that the tippers have a reason not to tip because of what that act would symbolise about their characters and commitments seems question-begging. If it would be morally wrong for them to act in this way, then doing so might well express a vicious character. But if there is no independent reason for them to refrain, then it is

difficult to see how tipping could express any untoward commitment or character trait.

Let's turn now to the version of the fairness approach defended by Cullity. On this view, the tippers have a collective obligation not to dump their waste in the reservoir, since that plurality of actions would jointly cause harm to the villagers. Moreover, when enough of the tippers refrain that the villagers' poisoning is averted, the others have a reason to refrain as well: since, under such conditions, tipping would be unfair. The difficulty facing this view was that if someone's actions are superfluous with respect to some cooperative effort, then it seems that it could not be unfair for them to sit out of that effort: even if they benefit from it. But if the fact that an action makes no difference to some outcome implies that it is superfluous with respect to it, then this causes problems for Cullity's approach; it implies that it could not be unfair for the tippers to tip, even if others refrain. Nefsky's defusal of the superfluity problem provides a way out of this difficulty. On her account, whilst no waste-tipping makes a difference to the harm suffered by the villagers, each can play a non-superfluous causal role in bringing it about. Therefore, if sufficiently many others refrain, it could be unfair for a given tipper to dump their waste.

Again, Nefsky's account seems able to steer Cullity's view around the superfluity problem. However, again, problems remain. Recall the rationale behind Cullity's appeal to the notion of collective obligation. This notion, I suggested, enables him to get around the second difficulty which faces fairness views: that considerations of fairness give us conditional reasons for action. The idea is that in instances where the tippers do not have a fairness-based reason, the collective obligation steps in to fill the gap. However, there is a problem with this strategy.

At the outset of Section 3.3, I claimed that the fairness approach aims to resolve the paradox of collective harm by denying clause (b) of the harmful/harmless series premise; in cases like that of the reservoir, the individuals in question each have an action-, rather than an outcome-, reason

to refrain from one of a series of collectively harmful/individually harmless actions. In instances where the tippers each have a fairness-based reason not to tip, this result follows. However, in instances where reasons of fairness are substituted for a collective obligation not to tip, this inference no longer goes through.

To illustrate, the sentence 'the soldiers surrounded the castle' would not usually be taken imply that any one soldier surrounded any castle; here, the property of surrounding the castle is attributed to the soldiers non-distributively. Similarly, I am suggesting that, where the obligation in question is collective, the sentence 'these individuals are obligated to φ' should not be taken to imply that any one of those individuals bears an obligation to φ. But if so, then the assumption that the tippers are collectively obligated not to tip will not imply that any one of them has a reason not to tip. As such, that assumption will not help to bolster the solution to the paradox proposed by the fairness approach.

This follows from the way in which I construed the notion of collective obligation at the outset of Section 3.3.2. But, one might object, why couldn't we just construe this concept in a way which avoids the problem? The alternative would be to say that, where the obligation in question is collective, the sentence 'these individuals are obligated to φ' ascribes the property of being obligated to φ of the relevant plurality distributively. This would imply that each one of the individuals is obligated to φ, and so would get us around the problem just described. However, interpreted in this way, the assumption that individuals bear a collective obligation to refrain from actions which together cause harm is question begging. That assumption entails the collective harm premise: the claim that if an action is one of several which will collectively cause harm to others, then its agent must have a reason to refrain from it. But in order to assume this, we need an argument to show that at least one of the other components of the paradox is false, an argument which remains forthcoming.

Of course, if we endorse the non-distributive reading of collective obligation assumed here, it does not follow that we could never derive individual obligations from a collective obligation; we would just need some further assumptions. Indeed, this is precisely how Cullity's invocation of fairness seems to function. This says that when a plurality of agents *xx* are collectively obligated to produce some outcome *o*, and sufficiently many of *xx* act so as to produce *o*, then it is unfair for others among *xx* not to contribute to that effort. However, when these further premises do not hold, the fact that several individuals bear a collective obligation gives us no grounds for claiming that any one of them has a reason to act in a certain way.

3.5. Collective reasons

I want to consider one final attempt to resolve the paradox of collective harm. One might wonder whether my comments at the end of the preceding section didn't misconceive the point of an appeal to the notion of collective obligation: 'The point' one might press 'is not to show that clause (b) of the harmful/harmless series premise is false. Rather, the point is to show that the collective harm premise is false. It is a mistake to think that if an action is one of several which will collectively cause harm to others, then its agent must have a reason not to perform it. However, the fact that several actions will be collectively harmful does constitute a *collective reason* to avoid them: that is, a reason which counts against those actions, without thereby counting against any one of them.'[109]

On this basis, we might replace the collective harm premise with the following variant:

The new-and-improved collective harm premise: If several actions will collectively cause harm to others, then the individuals in question must have a collective reason not to perform them.

[109] The idea of a collective reason for action, or something close to it, has explored independently by Alexander Dietz (2016), Frank Jackson (1987), and Thomas Smith (2009).

Since the claim that several individuals have a collective reason to perform/avoid some combination of actions does not imply that they each have a singular reason to perform/avoid some action, this variant is consistent with the conjunction of the inefficacy premise and the harmful/harmless series premise. Thus, no paradox.

This does not yet resolve the problem, however. In addition, it must be shown that our grounds for endorsing the new-and-improved collective harm premise are at least as independently compelling as those for endorsing the original. The grounds for endorsing the original, remember, are abductive: harm caused in cases like that of the reservoir seems to be morally objectionable; the best explanation of this datum is that actions which collectively cause harm are morally impermissible; and this implies the collective harm premise. Can the new-and-improved variant be defended on similar grounds?

The best case for the new-and-improved collective harm premise seems to be this. If, in the reservoir case, the villagers are harmed, the explanation of its objectionability is not that each tipper was obligated to refrain from tipping. Rather, it is objectionable because, together, the tippers bear a collective obligation to refrain from that harmful combination of actions. That is, whilst a given tipper would not violate any of their singular moral obligations by dumping their waste into the reservoir, the plurality consisting of all of the tippers would contravene their collective obligation if any one of them were to do so. But if we explain the objectionability of the villager's poisoning in this way, then the new-and-improved collective harm premise must be true; that is, the fact that several actions jointly cause harm must constitute, or imply, a collective reason to refrain from them.

I have some sympathy for this sort of approach. Indeed, as we will see in the following chapter, I agree that the best way in which to resolve the paradox of collective harm is to reject the collective harm premise. However, as presented above, the collective reasons approach faces a number of difficulties.

One initial difficulty can, I think, be dealt with. 'Why think' one might object 'that *all* of the tippers have a collective reason not to dump their waste into the reservoir? This plurality is just as capable of causing harm as each of those which comprise of all of the tippers minus one: since a single waste-tipping makes no difference. As such, the grounds for identifying all of the tippers as the relevant plurality seem just as good as those for identifying it with any of these smaller pluralities.

'However, if we concede this much, then we concede that each one of the tippers can reasonably deny that they belong to a plurality who have a collective reason not to tip. Each can truthfully say that the plurality containing every tipper but themselves would cause just as much harm as that containing all of them, if every member chose to tip. So, why not identify the former, smaller plurality as the ones who have the relevant collective reason? But if we allow this, then every tipper will have reasonable grounds for denying that they are one of several individuals who have a collective reason not to dump their waste.'

This objection assumes that there must be only one plurality of tippers who have a collective reason not to dump waste into the reservoir. But why think this? Why not think, instead, that there are many overlapping pluralities which each have such a reason? On this proposal, every plurality of tippers who could, if they all tipped, jointly cause harm to the villagers have a collective reason not to do so. As such, every tipper will belong to several such pluralities.

There is, however, a deeper difficulty with the collective reasons approach. It pertains to the very intelligibility of the notion of a collective reason (and so, by association, to that of a collective obligation as well). Up until now, I have assumed that we can make sense of this idea. But can we?

Here is one way in which to grasp the difficulty. The notion of a reason for action is logically tied to that of acting for a reason. Quite what that

connection is is a delicate matter, but it is something like the following. Suppose an agent reflects that they have a reason to φ, and that they have no reason not to φ. On this basis, it seems appropriate for the agent to conclude 'So, I shall φ!', and subsequently for them to φ. Moreover, when they φ, they will be acting for a reason: they will be φ-ing on the basis of the reason which they apprehended in practical thought.

However, if we suppose that there are collective reasons for action, the connection between this sort of reason and rational action becomes more obscure. Suppose that two agents share a collective reason to enact some combination of actions, and that they have no reason (singular or collective) not to enact that combination. Each of them reflects that this is the case. How are they to proceed? Neither one of the agents can enact the action-combination which the collective reason calls for: since this combination comprises the actions of two agents, and neither agent cannot perform the other's action. Furthermore, as we have seen, the claim that these two share a collective reason does not, by itself, imply that either one of them has a singular reason to do something. So how are these agents to end their state of inertia? How do they get from apprehending their collective reason, to acting for a reason?

For now, these too-brief comments are the best I can do to motivate scepticism about the idea of a collective reason. They do not render it nonsensical. But they do, I think, indicate a lacuna in accounts of the notion; and if that lacuna cannot be filled, then we will have good grounds for suspicion.[110]

A proponent of the collective reason approach might make the following reply: 'Admittedly, the notion of a collective reason is strange. It is unlike that of a singular reason, in that an individual cannot always employ the

[110] One possible way in which to elaborate the connection between collective reasons and rational action might be to draw on the game-theoretic notion of 'team reasoning', as developed by Michael Bacharach (2006) and Robert Sugden (1993). A discussion of this concept would, however, take me too far away from my topic.

premise that they have such a reason in a chain of reasoning which terminates in an intentional action. But the paradox of collective harm demonstrates that we must embrace such notions. Indeed, this paradox arises precisely because the prevailing conception of practical rationality is overly individualistic.' This response, however, presumes that there is no other way in which to solve the problem. The following chapter presents an alternative.

5

A Solution

The preceding chapter described a problem, and then proceeded to reject a number of potential solutions to it. My aim in this final chapter is to propose a novel solution to that problem. Section 1 offers a refresher on the problem detailed in Chapter 4, and explains how the solution to be developed in the proceeding sections differs from those previously considered. Section 2 goes on to offer a solution which relies on the assumption that there is a moral obligation to obey the law, whilst Section 3 extends the account by dropping that assumption. In Section 4, I contrast the solution given in Sections 2 and 3 with two similar accounts in the literature. Section 5 closes by considering some objections.

1. A fresh start

The problem introduced in Chapter 4 was that three seemingly plausible claims are incompatible:

The paradox of collective harm

(i) ***The collective harm premise:*** If an action is one of several which will collectively cause harm to others, then one must have a reason not to perform it.

(ii) ***The inefficacy premise:*** One has no outcome-reason to refrain from an action if, considered by itself, it is harmless.

(iii) ***The harmful/harmless series premise:*** It is possible:

(a) for a series of actions to be collectively harmful, whilst each member of that series is individually harmless; and

(b) for there to be such a series such that if the relevant individuals have a reason to refrain from their collectively harmful actions, then it must be an outcome-reason.

The case for (i), recall, was abductive: given certain background assumptions, it is intuitive to consider collectively caused harm to be morally objectionable; this is best explained by supposing that the individuals in question are morally obligated to refrain from their collectively harmful actions; and that supposition implies the collective harm premise. The case for (ii) stemmed from the notion of an outcome-reason, and a broad conception of harm: if we assume a conception of harm broad enough to cover any of an action's bad consequences, then it is difficult to see how someone could have a reason to refrain from a harmless action which is grounded in the badness of that action's outcomes. And the case for (iii) derived from the possibility of cases like the following:

The Reservoir: A large number of individuals (the tippers) each possess equal amounts of industrial waste of which they wish to dispose. When consumed in small quantities, the waste is completely harmless. However, if consumed in larger quantities, it is toxic. A potential dumping ground is a local reservoir, on the banks of which lies a hamlet whose inhabitants use the reservoir as their main source of drinking water. The quantity of waste possessed by each of the tippers is so small that, regardless of how much waste is already in the reservoir, if any one of them were to add their share to its waters, it would make absolutely no difference to the villagers' health. However, if all, or sufficiently many, of the tippers were to do so, then the villagers would be poisoned as a joint result.

And yet, taken together, (i)-(iii) imply a contradiction:

(iv) An individual can both have a reason and have no reason to refrain from performing one of a series of collectively harmful/individually harmless actions.

So, at least one of the components of the paradox must be false. But which? The preceding chapter explored five different attempts to resolve the paradox of collective harm. With the exception of the collective reasons approach, all of these proposed a solution which allowed us to retain the collective harm

premise; either (ii), (iiia), or (iiib) was singled out as the claim to be rejected. However, each solution was ultimately found to be untenable. Perhaps, then, it is time for a fresh start.

In this chapter, I offer a solution to the paradox of collective harm which breaks sharply from the leading approaches. The dominant view is that a successful solution should tell us where the thought 'Things will be just bad whether or not I act in this way, so I have no reason to refrain' goes wrong. Some authors locate the problem in the major premise; others argue that the inference itself is invalid.[111] Yet most are agreed on the rough shape which a solution ought to take. The solution developed in this chapter, however, resists that trend. I will argue that we should resolve the paradox of collective harm by rejecting the collective harm premise. If I am right, then it is entirely possible for each of a plurality of individuals to lack any reason to refrain from their respective actions, even if, taken together, those actions cause harm to others.

We should reject the collective harm premise, I propose, because its abductive rationale is flawed. I will argue that we can account for the objectionability of collectively caused harm in a way which does not entail the collective harm premise. According to this rival explanation, where collectively caused harm is objectionable, this need not be because the individuals whose actions caused it transgressed a moral obligation. Rather, it could be because some further individuals were obligated to give them reasons not to act in a collectively harmful way, and yet failed to do so. Other things being equal, this account is preferable to that which supposes that we are obligated not to do things which together cause harm to others; the reason for this being that the rival account does not imply the collective harm premise, and thus avoids paradox.

2. Law and authority

[111] Julia Nefsky (2015), for instance, sets out the possible strategies for resolving the problem in this way.

My alternative account of the moral objection to collective harm draws on the idea that, as rational agents, we are capable not only of appreciating and acting on the reasons we already have, but also of *creating* new ones. By this, I mean simply that by doing certain things, an individual can alter the reasons which she and others can act on. Clear examples are provided by the so-called 'moral powers': powers which enable an individual to impose or suspend an obligation. For instance, by promising to do something, a person can give herself an obligation to do it. However, there are many other examples which do not involve the imposition or suspension of any obligation: by offering you money in return for a lift, I can give you a reason to drive me; by placing a tack on your chair, I can give you a reason not to sit on it; by threatening to reveal your secrets, I can give you a reason to comply with my demands.

The legal system is a social institution which enables particular individuals to manipulate their own and others' reasons for action on a large scale. By making changes to the existing system of legal regulations, legislators can give those within their jurisdiction reasons to perform/avoid certain sorts of action which, but for those legislative interventions, they would otherwise lack. The means by which the law can alter people's reasons for action are various. For instance, if legislators criminalise an act, then individuals will be liable to be punished if they engage in it. Alternatively, if that act is subject to civil regulation, then individuals who enact tokens of it may be liable to pay a fine, or to compensate individuals who are harmed as a result. Threats of criminal punishment, non-punitive sanctions, and compensatory duties can all give individuals reasons to comply with legal regulations, reasons which, but for those threats, they would otherwise lack. Lawgivers can also manipulate people's reasons for action by making offers. For instance, a policymaker might propose a tax break for married couples. By doing so, she gives non-married individuals a reason to marry by offering them a benefit in return.

However, the law's capacity to create reasons for action may go beyond its power to threaten and offer. Some theorists have thought that legal directives themselves can provide those subject to them with reasons to act, reasons

independent of the costs attached to non-compliance.[112] More than this, some have thought that legal directives can be *morally binding*. On this view, not only can a legal directive give an individual a reason to act as directed; it can morally obligate them to do so. If so, then lawgivers possess a certain sort of moral power: by enacting changes to the directives made by the legal system, they can alter what those subject to those directives are morally permitted to do. I will refer to this as the view that the law possesses *authority*, or is *authoritative*.

2.1. A rival account

The capacity of the legal system to manipulate our reasons for action, I propose, can be drawn on to provide an explanation of the objectionability of collectively caused harm which does not turn on the collective harm premise. To develop this, let's return to the reservoir case. For the time being, I am going to introduce two new assumptions into the description of this example. These assumptions will be stripped away later on. For now, however, I will assume that they hold. First, let's suppose that the tippers and the villagers all fall within the jurisdiction of a common legal system. And second, let's suppose that the law is authoritative: that those subject to a justified legal directive have a moral obligation to comply with it.[113]

These two assumptions enable us to explain why the villagers have a claim against being harmed as a joint result of the tippers' actions in the following way. Suppose that legislators have made it illegal for the tippers to dump their waste into the reservoir: either by criminalising those actions, or by imposing

[112] As I understand the term, a 'legal directive' could belong to either the criminal or the civil law. Criminalisation is commonly understood as prohibition; when an act is criminalised, individuals are directed not to engage in it. However, civil regulations can also be so understood. A stipulation that individuals who act in a certain way are liable to pay a fine, or damages, can be interpreted as a prohibition on that act. See (Tadros 2010, pp. 167-169). Antony Duff (2002) has argued that we should think of the criminal law typically as *declaring* the prior wrongfulness of the acts defined as criminal, rather than as *prohibiting* them. The notion of prohibition, he thinks, should be reserved for describing criminal regulations that give individuals reasons to avoid behaviour which they would otherwise have no reason to avoid. Since I will be solely concerned with regulations of this latter sort, I will retain the terms 'prohibition' and 'proscription', which I will use interchangeably.

[113] I will not rehearse the reasons why one might think that this claim is true. For a recent defence, see (Scheffler 2018). For an extensive critical survey, see (Simmons 1979).

civil regulations on them. If this legal directive is justifiable, then, given the law's authority, it follows that the tippers are morally obligated not to tip. But recall the assumption I made in Section 2.1 of Chapter 4: harm is objectionable if someone was obligated to take certain precautions to prevent it, and yet failed to take them. Given this, it follows that if the tippers were to transgress the law against waste-tipping and dump their waste in the reservoir regardless, then the harm their actions would jointly cause would be morally objectionable: since they would be obligated not to tip their waste. Moreover, this does not imply the collective harm premise; but for this law, the tippers might have no reason whatsoever to refrain from dumping their waste into the reservoir.

'Fine,' one might respond 'but what if there were no legal directive against tipping waste in the reservoir? Under such conditions, it seems as though the villagers would still have a valid claim against being poisoned as a joint result of the tippers' actions. Yet this cannot be accounted for by the existence of a legal directive which makes it wrong for the tippers to tip: since there is no such directive. How, then, can we explain our datum without saying something which implies the collective harm premise?'

I reply as follows. Notice that, in this case, the legislators are in a position to protect the villagers from the harm which the tippers' actions stand collectively to cause them. Assuming that people are motivated not to act wrongfully, if it were (justifiably) made illegal for the tippers to dump their waste, then would they seek to avoid those actions: since, given the law's authority, they would be morally obligated to do so. So, by proscribing acts of waste-tipping, the legislators could pre-empt the harm which the tippers might otherwise jointly cause.

Typically, when a person is in a position to protect another from harm, we tend to think that they ought to do so; we think, that is, that we bear a duty of assistance which can obligate us to save others from avoidable harm. There are, of course, exceptions to this rule of thumb. If by providing you my protection I would incur very great costs, then it might be admirable, but not

obligatory, for me to do so. Or perhaps my help would cause a great deal of harm to some third party, and as such would, on balance, be morally objectionable. Or maybe the harm you stand to suffer is the foreseen result of a series of choices which you have soberly made, meaning that others do not owe you their assistance. However, in the absence of such extenuating factors, we tend to think that we are morally obligated to extend such help to one another.

Given all of this, it seems we can say that the legislators in the reservoir case are obligated to introduce legal directives which prohibit the tippers from dumping their waste in the reservoir. By doing so, they could protect the villagers from avoidable harm. Furthermore, the extenuating circumstances which might exempt these individuals from their duty of assistance seem to be absent; implementing these laws, we can assume, would not involve any great sacrifices on their part, nor would they impose any significant costs on any third party. As such, the legislators bear an obligation to take steps to prevent the villagers from being harmed by the tippers' actions, in the form of implementing legal proscriptions on such behaviour. So, if they do not take these steps, and the villagers are subsequently harmed, then the harm they suffer will be morally objectionable.

If all of this is correct, then the collective harm premise will begin to look under-motivated. In the event that a legal directive against acts of waste-tipping is implemented, we can explain the objectionability of any harm caused by such actions without appealing to that premise. The objection is that the tippers were obligated not to tip. But this obligation derives from the authority of law; but for a legal prohibition on such actions, the tippers might have no reason whatsoever to refrain from dumping their waste into the reservoir. Moreover, when there is no such prohibition, the objectionability of any harm jointly caused by acts of waste-tipping can again be accounted for without appeal to the collective harm premise. Here, however, the objection is not that the tippers' actions are morally impermissible. Rather, it is that the legislators' *inaction* is unjustifiable; these individuals could have protected the villagers, yet failed to do so.

As I have already made clear, this rival account of the moral objection to collective harm pivots on the two assumptions I introduced at the outset of this subsection: that the tippers and villagers are all subject to a common legal system, and that the law is authoritative. These assumptions will be addressed in due course. However, the account also relies on a further assumption which, up until now, I have not explicitly discussed: namely, that it is justifiable for legislators to proscribe the tippers' acts of waste-tipping.

This third assumption is required in order for a legal directive against waste-tipping to be morally binding. Even if we suppose that the law is authoritative, it does not follow that those subject to it are morally obligated to conform to just any law which their government chooses to introduce. We would not, for instance, be obligated to obey a law which granted the use of public parks only to those whose names contained an odd number of letters. In order to be binding, a law must be justifiable: there must, that is, be sufficient reason for legislators to implement it. As such, if my rival account is going to work, there must be sufficient reason for lawgivers to prohibit collectively harmful/individually harmless actions; and this must be so, moreover, even if, but for that prohibition, there would be no reason for anyone to avoid them. Demonstrating the truth of this assumption is the task of the following subsection.

2.2. Justifying prohibition

What reason is there for legislators to prohibit actions which, whilst individually harmless, together cause harm? There is an obvious possible answer. As explained above, given the law's authority, such a prohibition could pre-empt the harm which those actions might otherwise cause. That is, prohibiting collectively harmful/individually harmless actions is likely to protect individuals from harm which others might do to them.

Typically, if a legal directive would protect individuals from being harmed by others, then that is a good reason to implement it. There are, however, complications. Some of these need not concern us here. For instance, there

might be certain types of harm which a state cannot justifiably claim the prerogative to protect its citizens from: harms to an individual caused by actions which fall within her own legitimate sphere of control, perhaps (Shiffrin 2000); or certain harms which eventuate from beliefs which a person adopts on the basis of others' speech (Scanlon 1972). But the harms which stand to be suffered in the cases which here concern us, we can assume, are not of this type; these are harms which a state can justifiably protect its members from.

However, there are other difficulties which are relevant to the present discussion. The following subsections address two of these difficulties.

2.2.1. More harm than good?

The fact that a legal directive would prevent certain individuals from being harmed by others is not, by itself, a sufficient reason to implement it. Implementing the directive may impose harm on some further individuals: for instance, by restricting people's freedom, or by virtue of the costs of enforcing it. If these harms outweigh the harm which the directive prevents, then, on balance, legislators may have good reason not the implement it.

When dealing with toy examples, such as the reservoir, we can simply assume that the costs of proscribing the relevant collectively harmful behaviour are outweighed by the harm which doing so would prevent. However, as we approach the real world, this assumption can be called into question. As I noted in Chapter 3 (Section 2.3), there will be a variety of cases in which a legal prohibition on a certain sort of collectively harmful behaviour would likely do more harm than good. And yet, it seems we do not want to deny that the individuals who stand to be harmed in such cases have a claim against being harmed in that way. How, then, can we account for that datum?

We can explain it, I suggest, if we suppose that, where a prohibition on collectively harmful behaviour would be excessively costly, legislators are obligated either to prevent or to mitigate the harm which stands to be caused by other means. For instance, an alteration to the taxation system might

incentivise individuals not to act in collectively harmful ways, without prohibiting them from doing so. Alternatively, changes in other areas of policy might pacify behaviour which would otherwise have harmful cumulative effects (think, for instance, of the legislative provisions to tackle gender inequality mentioned at the end of Chapter 3's Section 2.3). However, to keep things simple, I will concentrate on cases where proscribing collectively harmful behaviour would prevent harm on balance. Non-prohibitive regulation will be taken up again in Section 3.

2.2.2. The dependence thesis

Let's suppose, then, that, by implementing a legal directive against some collectively harmful act, legislators could protect certain individuals from being harmed by others, without imposing countervailing costs somewhere else. Do they therefore have a sufficient reason to implement it?

One might think that the answer is clearly 'yes'; and, in fact, I think that this is the correct answer. However, things get complicated when we suppose, as I suggest we should, that the collective harm premise is false. This claim, remember, says that if an action is one of several which will collectively cause harm to others, then its agent must have a reason to refrain from it. If this claim is false, then several individuals could each lack any reason to refrain from their respective actions, even if, taken together, those actions cause harm to others. Could a legislator have sufficient reason to prohibit a piece of collectively harmful behaviour if, but for that prohibition, individuals would have no reason not to engage in it?

Our answer may differ depending on how the regulation in question is enforced. If we have no pre-legal reason to avoid some collectively harmful act, then perhaps it could be justifiable to regulate it through the civil law: that is, by stipulating that one is liable to pay damages, or some purely administrative fine, if one engages in the prohibited act. This proposal is not without its problems. For instance, if an action is singly harmless, then what damage could its agent be made liable to compensate? But such difficulties are not insurmountable. For instance, individuals who enact a plurality of

collectively harmful/singly harmless actions could *each* be made liable to pay the *total* cost of the harm jointly caused by the plurality. Alternatively, each could be made liable to pay the total cost of the *excess* harm: that is, the harm which exceeds that which could be expected if all of the individuals in question were to follow the optimal course of action.[114]

However, one might argue that an act which we have no pre-legal reason to avoid cannot justifiably be criminalised. In order for it be justifiable for legislators to criminalise an act, one might think, that act must be wrongful, or at least seriously harmful, independent of its definition as a criminal offence. Yet, if the collective harm premise is false, then it will be possible for an action to cause harm jointly with others without being wrongful (absent its being defined as a criminal offence). This simple conception of the proper scope of the criminal law seems doubtful. For one thing, it would render large swathes of existing criminal law unjustifiable: specifically, so-called *mala prohibita*: that is, criminal offences which are wrongful solely in virtue of the fact that they have been criminalised.[115] However, for present purposes, I propose to put questions about how regulations on collectively harmful behaviour ought to be enforced to one side. Whichever way we decide to answer such questions, there is a more basic difficulty facing my account.

The problem is this. I want to propose that it can be justifiable for legislators to prohibit collectively harmful acts in order to prevent harm, even if individuals have no pre-legal reason to avoid such acts. However, this claim creates tension with an influential account of legal authority: namely, Joseph Raz's *service conception*. According to this view, the primary function of a political authority (a government with the right to issue morally binding laws) is to help its subjects to comply with the reasons which apply to them independently of its directives (Raz 1986, chap. 3, esp. pp. 55-56). In order

[114] For the second proposal, see (Cooter and Porat 2007). The authors note, however, that a rule of total liability for excessive harm would be impractical in cases where harm is caused jointly by very large numbers of individuals. For a general discussion of compensatory schemes which differ from schemes of tort liability, see (Tadros 2010, pp. 178-181).
[115] For a helpful discussion of how *mala prohibita* may be brought within the proper scope of the criminal law, see (Duff 2002).

to be justifiable, Raz claims that authoritative directives must satisfy the *dependence thesis*: "all authoritative directives should be based on reasons which already independently apply to the subjects of the directives and are relevant to their action in the circumstances covered by the directive" (Raz 1986, p. 47).

The tension seems to be this. I am proposing that we do not have a reason to refrain from a series of actions in virtue of the fact that they together cause harm. However, I am also proposing that legislators can be justified in proscribing our actions on the grounds that they are collectively harmful. As such, what I am claiming is that a prohibition on collectively harmful behaviour can be justified by a consideration which does not itself imply any reason for individuals to avoid the actions which that directive prohibits. But this seems to be in direct contravention of the dependence thesis. How could such a law help individuals to act on reasons which they already have, independently of those which the law itself introduces?

This tension is not fatal for my approach. Even if the truth of dependence thesis and the falsity of the collective harm premise jointly imply that it is largely unjustifiable for lawgivers to prohibit collectively harmful behaviour, there may still be other, non-prohibitive measures which they could take to protect people from the cumulative effects of such behaviour. However, it would severely limit my account. One way out would be simply to reject the service conception of authority. This move, however, is unnecessary. I submit that, on its most plausible interpretation, Raz's dependence thesis is in fact consistent with my proposed justification for prohibitions on collectively harmful/individually harmless actions.

The dependence thesis, I suggest, cannot be interpreted as requiring that authoritative laws be based exclusively on reasons which we already have to act as they direct, and would otherwise have even in the absence of those laws. If it did require this, then the dependence thesis would imply that authoritative legal directives are unable to alter what we have reason to do. However, Raz argues that it is a major advantage of the service conception

over rival conceptions of authority that it does *not* have this implication. Any view which implies that authoritative directives do not change our reasons for action, he claims, will be unable to account for some of the central functions of authority (Raz 1986, p. 30, 48-51). For instance, it will be unable to account for the role of authorities in resolving coordination problems; in order to do this, authorities must be able not only to tell us what we already have reason to do, but also to *give* us reasons to act.[116]

So as to prevent Raz's conception of authority from falling foul of his own criticisms, I propose that the dependence thesis should be interpreted as follows: in order for a law to be authoritative, those subject to it must typically have a further reason to obey it beyond the very fact that it is the law, and the prudential reasons introduced by any penalties attached to non-conformity. On this interpretation, the thesis does not imply that a law lacks authority if, *prior to* its implementation, individuals lacked any reason to act as it directs: just so long as, *once it has been introduced*, those subject to the legal directive typically have an independent reason to comply with it.

This interpretation of the dependence thesis, I suggest, is consistent with my justification of legal proscriptions on collectively harmful/individually harmless actions. The proposed reason for introducing such proscriptions is that they can protect individuals from harm which others might otherwise do to them. When an effort to protect people from serious harm is made, others have a reason not to compromise it: that is, not to do things which will lower its chances of success. By transgressing a prohibition on collectively harmful/individually harmless actions, I suggest that one can compromise that law's chances of protecting people from harm. As such, we have an independent reason to comply with such prohibitions.

[116] A. J. Julius (2013) claims that it is never justifiable to give someone a reason to do something if they would otherwise lack a reason to do it. Raz, I am proposing, gives us grounds to doubt this claim. For further objections to Julius' view, see (Kolodny 2017; White 2017).

But, one might press, how exactly can an act of transgression compromise legislators' efforts to protect others? It cannot be that, by disobeying a prohibition on collectively harmful/individually harmless acts, an individual causes harm to others; by stipulation, such acts are singly harmless. Nor can it be that people generally have a sheeplike disposition to imitate the actions of others. This assumption would give us a much simpler solution to the paradox of collective harm: we have a reason to avoid collectively harmful/individually harmless actions, since by performing one we will cause others to follow suit, thereby causing harm indirectly.[117] How, then, does my proposal work?

Notice that my proposed justification for prohibitions on collectively harmful/individually harmless acts relies on the idea that legal directives have the power to influence how people behave. In order for lawgivers to be able to use the law as a means to protect people from harm which others might do to them, it must have an influence over how people deliberate and act; if laws had no such influence, then they would be powerless to prevent people from acting in harmful ways. Here, I am assuming that the power of the law to influence behaviour stems from its authority: that is, from the fact that legal directives can be morally binding. However, the authority of a given law is not sufficient to endow it with the power to influence. In addition, those subject to it must *believe* that they are morally bound to obey. If they lack this belief, then the mere fact of the law's authority will not induce compliance.

What I want to suggest is that by transgressing a prohibition on collectively harmful/individually harmless actions, an individual can give others grounds to doubt that they themselves are obligated to comply with that particular prohibition. If so, then that act can undermine the prohibition's power to influence behaviour, and thereby compromise its chances of protecting people from harm.

[117] For objections to a solution along these lines, see (Nefsky 2018, pp. 271-272; Sinnott-Armstrong 2005, pp. 299-300).

Given my proposed justification for prohibitions on collectively harmful/individually harmless acts, individuals will be morally bound to comply with such prohibitions only if sufficiently many others also comply. If enough individuals transgress such a prohibition that their actions jointly cause harm to others, then it will be false that that prohibition prevents others from being harmed. It will therefore be unjustified, meaning that those subject to it will no longer be morally bound to comply with it: or at least, they won't be bound by virtue of its authority. This implies that individuals will have reason to believe that they are bound by a given proscription on collectively harmful/individually harmless behaviour only if they can be assured that sufficiently many others will comply with it. As soon as this assurance is lost, these individuals have reason to doubt that they are obligated to obey the directive.[118]

However, by transgressing a law which prohibits collectively harmful/individually harmless behaviour, an individual can, given certain background conditions, undermine others' assurance of general compliance. Such an action can give others reason to suspect that sufficiently many individuals might act likewise. This will be the case, for instance, if these individuals have reason to think that the transgressor's behaviour is explained by some motivational or cognitive state (strong countervailing motivations, lack of respect for this particular law, lack of knowledge about the legal situation etc.) which is shared by sufficiently many others.

So, if a single act of transgression can reasonably be taken by others to reveal a common disposition towards non-compliance, then it follows that it can give those others reason to doubt that they themselves are morally bound to

[118] This might seem to undermine my earlier claim that when tipping is proscribed by a morally binding directive, the moral objection to the villagers' poisoning can be explained in terms of the tippers' wrongful disobedience to the law: if the villagers are harmed, then the directive will have failed to protect them, and so will cease to be binding. However, this is too quick. The claim I am putting forward is this: if the level of compliance with a directive is so low that continuing to enforce it will not prevent harm, then that directive may cease to be binding. But this does not imply that the directive in question was *never* morally binding. Those who originally transgressed it may have thereby acted wrongfully, even if their noncompliance brings it about that, later down the line, others can permissibly transgress the directive.

comply with the legal directive in question. This in turn can strip that directive of its power to influence behaviour by means of the law's authority.[119] And this, finally, compromises its effectiveness as a means to protect people from others' harmful behaviour. As such, given the right background conditions, individuals have an independent reason to comply with prohibitions on collectively harmful/individually harmless acts: even if the collective harm premise is false.

It should be noted, however, that this reason will not always apply. An act of disobedience will not always compromise the chances which such a prohibition has of preventing harm. Whether or not it does will depend on a number of factors: how public the transgression is; whether there are grounds to suspect that others will follow suit etc. For Raz, however, this does not matter. The service conception does not require that individuals *always* have an independent reason to comply with authoritative directives. Rather, it requires merely that individuals will be *more likely* to conform to such reasons if they act as an authority directs them to, compared with if they were to judge the merits of each case for themselves (Raz 1986, pp. 55-56). So long as it is less likely that a given individual will compromise an effort to prevent harm by simply complying with a prohibition on collectively harmful behaviour, compared with if she were to judge each case herself, she is, according to the service conception, morally bound to obey that prohibition: even if, on occasion, transgressing the prohibition would not cause her to compromise such an effort.

3. Law without authority

The foregoing section offered an account of the wrong of collective harm which assumed that the law is authoritative; it was assumed that the very fact of an act's illegality could obligate individuals not to enact it. But what if this assumption is mistaken? Could we nonetheless explain the objectionability of collectively caused harm without appealing to the collective harm premise?

[119] It might nevertheless retain the power to influence behaviour through other means: coercion, for instance. This point will be addressed in Section 3.1.

Section 2's account relies on the law's authority at two points. First, it was appealed to in order to account for the objectionability of harm caused jointly by *illegal*, individually harmless actions. Second, the authority of law was relied upon in order to establish that harm caused jointly by individually harmless actions could be objectionable, even if those actions were *permitted* by morally binding laws. In this section, I will argue that such harm can be morally unacceptable, even if (a) the collective harm premise is false, and (b) the law is not authoritative. Section 3.1 accounts for the objectionability of harm caused collectively by individually harmless actions which are *permitted* by non-binding laws. In Section 3.2, I argue that harm caused jointly by individually harmless actions can be objectionable even when those actions are *prohibited* by non-binding laws.[120]

3.1. Coercion and autonomy

In Section 2, I offered the following argument for thinking that harm produced collectively by legal, individually harmless actions can be morally objectionable. By introducing morally binding proscriptions on collectively harmful/individually harmless actions, lawgivers could protect individuals from the harm which such actions might otherwise cause them. This is because, given that justified laws are morally binding, and that individuals seek to comply with their moral obligations, people will tend to avoid collectively harmful/individually harmless actions if they are illegal. Since they can offer such protection, lawgivers are often morally obligated to introduce such laws; and by neglecting that obligation, they can render objectionable harm caused by the actions they ought to have prohibited.

Can the same conclusion be reached without assuming that the law is morally binding? Section 2's argument appeals to the law's authority to explain its

[120] One might question whether the law could be described as issuing 'prohibitions' if it is not morally binding. Perhaps non-binding regulations merely offer us a choice: we can refrain from the regulated act and avoid whatever costs are attached to it, or we can engage in that act and shoulder the costs. For simplicity of phrasing, I will continue to describe certain non-binding legal regulations as 'prohibitions' and 'proscriptions'. However, if one prefers, this lexical choice could be revised without altering the substance of the account.

capacity to influence behaviour. However, this capacity can be readily explained without such an appeal. We can have prudential reasons to comply with a legal directive because of the penalties attached to non-conformity. Alternatively, policymakers can incentivise certain forms of behaviour by attaching benefits to its performance; tax breaks, for instance, could be offered to individuals or corporations whose conduct meets certain standards.

Legislators can, then, give people reasons to avoid collectively harmful/individually harmless behaviour, even if the legal directives they issue are not authoritative. By threatening to penalise those who fail to comply with non-binding prohibitions on such behaviour, they can give individuals prudential reasons to refrain from it.[121] Furthermore, policymakers could create such prudential reasons without issuing prohibitive legislation. By offering benefits to those who meet certain behavioural standards, they can incentivise individuals to refrain from collectively harmful actions. As such, legislators can protect individuals from the harm which such behaviour might otherwise cause, even if they cannot issue morally binding prohibitions.

This might seem to resolve the issue. However, others materialise. A notoriously thorny problem arises when legislators threaten individuals in order to impel them to comply with the law. It can be morally wrong to control others' behaviour by threatening them; we call this *coercion*. Would it, therefore, be wrong for legislators to threaten the public in order to get them to refrain from performing collectively harmful, though individually harmless, actions? If it would, then legislators could not be morally obligated to issue such threats: since they would be obligated both to make and not to

[121] Again, there is the issue of what form these penalties should take: criminal punishment, administrative fines, or the payment of damages? Perhaps the assumption that the law is not authoritative affects this choice; perhaps it is unjustifiable to punish people for doing things which are neither pre-legally wrong, nor wrong solely in virtue of the fact that they are illegal. Notice, however, that this view has reformist implications: it would imply that if, as many theorists think, the law is not authoritative, then *mala prohibita* will be unpunishable. Perhaps this is right, perhaps it is not. I will take no stand on the matter.

make them.[122] My argument would therefore fall apart. But if it would not be wrong for lawgivers to threaten the public in this way, then why?

This question becomes all the more pertinent when we assume that the actions which the legislators are impelling people not to perform are morally permissible. Perhaps it can be permissible to use threats in order to prevent a person from acting wrongfully, or from acting contrary to particularly good reasons. But if we reject both the collective harm premise and the claim that the law is authoritative, then this will not be the case for prohibitions on collectively harmful/individually harmless actions; indeed, but for the penalties attached to non-compliance, individuals might have no reason whatsoever to comply with these prohibitions. How, then, could it be justifiable for lawgivers to use the threat of penalties in order induce compliance with such legislation?

The wrong of coercion has been explicated in a number of ways, and to discuss all of them here would constitute too much of a detour.[123] I will therefore focus on one particularly prominent view. In the remainder of this subsection, I will argue that even if we assume that coercion is objectionable for the reasons given by this theory, it can nevertheless be justifiable for legislators to coerce people into avoiding collectively harmful/individually harmless behaviour.

According to the view which I will assume, coercion is morally objectionable because it invades personal autonomy. In his landmark discussion, Raz (1986, pp. 369-372) characterises an autonomous person as one who is an author of their own life, one who fashions the greater part of its content and character through their own independent choices. For Raz, there are a number of conditions which must be in place if a person is to enjoy a high degree of personal autonomy. One of these is *independence*. Someone enjoys

[122] Perhaps one could be morally obligated both to do and not to do something. However, it is most problematic for my account if this claim is false, and so I will assume that it is so.
[123] For some recent discussions of the wrong of coercion see (Julius 2013; Kolodny 2017; Pallikkathayil 2011; Scanlon 2018, chap. 7; White 2017).

independence when their will is not subjected to that of another: that is, when others do not try to control what they choose to do, or to manipulate their decision-forming processes. On Raz's view, coercion invades autonomy by undermining independence: by coercively threatening another person, one imposes one's own will on theirs. For this reason, there is always a *pro tanto* objection to coercive behaviour (Raz 1986 p. 155; pp. 377-378).

It will help to clarify this view to consider an objection to it. Stephen White (2017, pp. 200-201) worries that autonomy-based views such as Raz's risk triviality. This is because they seem to define autonomy such that it is tautologous to say that coercion invades autonomy. For instance, one might think that, in claiming that coercion undermines the independence condition for autonomy, Raz is simply building it into his definition of autonomy that the absence of coercion is necessary for a fully autonomous life. But if so, then the claim that coercion undermines autonomy will be trivially true: it will simply follow from the stipulated definition of autonomy. Yet if this is correct, then won't it also be trivial to say that coercion is wrong because it invades autonomy?

Let's grant that, in specifying the independence condition for autonomy, Raz is simply stipulating that coercion invades autonomy. Even so, it does not follow that it is trivial to claim that this is what explains the moral objection to coercion. To see this, consider an example which I draw from another chapter of *The Morality of Freedom*. Raz argues that the value of artworks can be understood as deriving from the fact that they are constituents of what he calls "a life with art", a life which involves experiencing and contemplating artworks (Raz 1986, p. 201). Understood in this way, the value of artworks is intrinsic rather than instrumental. That is, their value does not derive their effects, such as the sensations which they cause in their audiences, but rather from the value of the complexes of which they are a part (viz. lives with art).

In this example, the claim that artworks are a constituent of a life with art is simply stipulated in the definition of the latter. So, if we assume that

definition, then this claim will be tautologous. However, this does not make it tautologous to say that the value of artworks derives from the fact that they are a constituent of a life with art. This is because the definition of 'a life with art' does not stipulate that such a life is intrinsically valuable; this is left open for philosophical discussion. As such, whilst the claim that a life with art necessarily contains artworks may be trivial, the view that artworks derive their value from the truth of that claim is not: the possibility that their value is instrumental, rather than intrinsic, is not ruled out.

A similar point can be made in the case of coercion. Let's assume that Raz's definition of autonomy stipulates that coercion invades autonomy. Given this, if we adopt that definition, then the claim that coercion invades autonomy will be tautologous. However, it does not follow from this that the claim that coercion is wrong because it invades autonomy is a tautology. If the value of autonomy is not assumed by its definition, then the wrongfulness of actions which prevent a person from living a fully autonomous life will be open for debate. So, even if it is trivially true that coercion undermines autonomy, the view that the wrong of coercion is explained by the truth of that claim need not be.

We now have an account of the objection to coercion: coercion is morally objectionable because it undermines independence, and so prevents the coercee from leading a fully autonomous life. However, for Raz, independence is not the only condition necessary for full autonomy. Another is *adequacy of options*. A person enjoys this when they maintain a sufficiently broad range of decent alternatives. This is a necessary condition for full autonomy, on Raz's view, because if the quality of the options one faces are bad, if most of them do not provide for one's basic needs, or do not allow one to pursue the projects and relationships around which one's life is based, then one is not truly the author of one's own life; one's choices will be determined not by one's own design, but by the struggle to maintain a life worth living (Raz 1986, p. 155; pp. 376-377).

Actions other than coercion can therefore invade people's autonomy by adversely affecting the quality of their options. For example, suppose an employer pays her employees a subsistence wage. The meagreness of this wage constricts the range of decent options open to the employees: if they continue to work for their employer, then they can only afford to meet the basic needs of themselves and their families; but if they leave, then they risk extended unemployment. Their autonomy is thereby stunted. However, the employer does not coerce her employees. She keeps their wages low not in order to control their behaviour, but to please her shareholders. Actions like this can invade people's autonomy to a greater extent than certain instances of coercion; someone who is coerced but nonetheless maintains a range of decent options might be more autonomous than an uncoerced person who lacks such alternatives. Moreover, it might turn out that the only effective way in which to prevent such actions might be to coerce people into refraining from them. The only way to protect against pervasively low wages, for instance, might be for governments to introduce a minimum wage, and to penalise employers who fail to meet it. In such cases, Raz (1986, p. 156; p. 419) argues, coercion can be justified; even though it always invades autonomy, it is justifiable to coerce others if, but for that coercion, people's autonomy would be even more severely restricted.

An action which, by itself, is harmless cannot adversely affect the quality of other people's options. One might therefore think that it cannot be justifiable to coerce individuals into avoiding such actions. This, however, is a mistake. If several individually harmless actions can collectively cause harm when performed in combination, then they might *jointly* worsen the options of others: even if each is *individually* incapable of doing so. By dumping their waste in the reservoir, for instance, the tippers collectively restrict the range of decent options open to the villagers: they could remain where they are and risk being poisoned by the contaminated water; or they could move, and leave behind relationships, professions etc. around which they have structured their lives. The tippers' actions thus jointly invade the villagers' autonomy; once

the reservoir becomes toxic, the quality of their options is worsened.[124] When several actions stand jointly to worsen the options of others, the only effective way in which to protect people's autonomy may be to issue coercive threats. For instance, threatening to penalise anyone who tips any quantity of waste into the reservoir, might be the best way in which to protect the villagers' autonomy. When this is so, such coercion will be justified: even if, considered individually, each of those actions poses no threat to the autonomy of others.

We can, therefore, account for the objectionability of harm caused collectively by legal, individually harmless actions without appealing either to the authority of law, or to the collective harm premise. Since it can protect individuals from serious harm, lawmakers can be obligated to introduce non-binding legislation which gives people prudential reasons to avoid collectively harmful/individually harmless behaviour. So, if they are obligated to take this precaution, but fail to do so, then any harm caused jointly by the actions which legislators ought to have regulated will therefore be morally objectionable.

3.2. Non-compliance

What about harm jointly caused by *illegal*, individually harmless actions? In Section 2, I explained the objectionability of such harm by appealing to the authority of the law: since those actions are illegal, and since the law is morally binding, individuals are morally obligated to avoid them; and since they are so obligated, any harm which those actions jointly cause is objectionable. But can we reach the same conclusion without assuming either the law's authority, or the truth of the collective harm premise?

If the law is not authoritative, then a legal directive could not *make* it morally wrong for people to engage in collectively harmful/individually harmless behaviour. Furthermore, if the collective harm premise is false, then

[124] For Raz (1986, p. 414), to be harmed is to have one's autonomy impeded. As such, all cases of collectively caused harm will be instances in which some individual's autonomy has been invaded. Seana Shiffrin (2012) also connects the notion of harm with that of autonomy, though in a looser way than Raz.

individuals could lack any pre-legal moral obligation to refrain from such behaviour. They might become so obligated by, for instance, making a promise; but otherwise, collectively harmful/individually harmless actions will be morally permissible.

Suppose, then, that some collectively harmful/individually harmless act is proscribed by a non-binding law, but that enough people contravene this proscription to cause harm to others. Granted the foregoing, we will not be able to explain the objectionability of this harm by pointing out that those whose actions were its joint cause ought to have taken the precaution of not acting as they did; in fact, in contravening the law and thereby producing the harm, they behaved perfectly permissibly. How, then, can we account for the intuition that others can have a claim against being harmed in this way?

I propose the following answer. In Section 3.1, I argued that legislators can be morally obligated to prohibit collectively harmful/individually harmless behaviour, even if their prohibitions are non-binding. This is because, once backed by threats of penalty, such non-binding prohibitions can nevertheless offer protection against the harm which the prohibited actions might otherwise cause. However, if prohibitions of this kind are issued, yet in fact fail to offer such protection, can we truly say that the officials in question have offloaded their obligations?

The answer depends upon whether, even though it is not perfectly effective, the current legislative strategy nonetheless represents the best way in which to protect people from harm. Possible alternatives are numerous. The legislators might threaten non-compliers with a heavier penalty. Or they might try to make compliance more feasible; for example, if tipping in the reservoir is banned, the number of safe disposal units nearby could be increased. An effort might also be made to improve public knowledge of the prohibition, the penalties attached to non-compliance, and the alternative courses of action which remain open to the public. Alternatively, policymakers could try implementing non-prohibitive legislation which incentivises the desired behaviour, rather than proscribing and penalised the

unwanted behaviour. These are just a few obvious measures which public officials might take to improve the effectiveness of legal regulations; imaginative social science is sure to offer many more.[125] Unless they are fully explored, officials who introduce ineffective prohibitions on collectively harmful/individually harmless behaviour fail to offload their obligation to protect the public.

Harm caused jointly by illegal, individually harmless actions can therefore be morally objectionable, even if the law is not authoritative, and the collective harm premise is false. When such harm is objectionable, this is not because those whose actions jointly caused it ought not to have behaved as they did. Rather, it is because lawmakers ought to have done more to deter them from behaving as such.

4. Comparisons

Our grounds for endorsing the collective harm premise are weaker than they at first appeared. Initially, it seemed that we needed this claim in order to explain why individuals can have a claim against being harmed as a joint result of others' individually harmless actions. However, I have argued that this is not so. We can instead explain our intuition, I propose, by appealing to the idea that lawgivers bear a moral obligation to use their legislative powers to protect people from the harmful cumulative results of such actions. If I am right, then the abductive case for the collective harm premise is undermined; we do not need to suppose that this claim is true in order to account for the objectionability of collectively caused harm. But if so, then we have a solution

[125] One component of the policymaker's toolkit which I have not discussed is the notion of 'nudging'. A 'nudge' is an intervention which is enacted on an agent's environment in order to get her to behave in a certain way (often, in a way which serves her own long-term interests, though a person could be 'nudged' for some other reason, such as to prevent harm to others). However, it does this without prohibiting the agent from doing anything, nor by attaching costs to any of her options. Rather, the nudge takes advantage of her pre-existing decision-making heuristics and biases in order to make certain options more salient to her. See (Sunstein and Thaler 2008). Policy which makes use of nudges differs from the sorts of regulation I have discussed in the text; nudging does not give people reasons to act in certain ways, but rather presents their options in a way which makes some of them seem more attractive. Nonetheless, I accept that nudging might be an effective way for policymakers to deter people from engaging in collectively harmful/individually harmless behaviour.

to the paradox of collective harm: we can safely reject the collective harm premise, and retain the other two components of the paradox.

I mentioned at the outset of this chapter that the solution for which I am arguing differs markedly from the dominant approaches. Most views attempt to locate some fault in the inference 'My act makes no difference, so I have no reason to refrain'. By contrast, my view embraces that inference; absent some further factor which gives them such a reason, it may be perfectly correct for an agent to infer that they have no reason to avoid an action from the fact that, considered singly, it causes no harm: even if it is one of several which together cause harm. However, whilst it breaks from the dominant approach, my account is not entirely novel. Other authors have appealed to the responsibilities of lawgivers in order to resolve problems which arise in cases of collective action. In this section, I draw out some similarities and differences between my account and two others: one due to Walter Sinnott-Armstrong, and another due to Robert Goodin.

Sinnott-Armstrong (2005) has argued that, whilst activities which cause the emission of greenhouse gasses (GHGs) have harmful cumulative effects, this does not necessarily imply that it is morally wrong for an individual to engage in them. For instance, this fact, he claims, does not make it wrong for someone to take a gas-guzzling SUV out for a Sunday joyride. This is because a single joyride will not emit enough GHGs to make any difference to global temperatures, and so will not cause anyone any harm by virtue of its emissions. However, Sinnott-Armstrong claims that governments should nonetheless regulate such activities, even if they are (pre-legally) morally permissible. "It is better" he writes "to enjoy your Sunday driving while working to change the law so as to make it illegal for you to enjoy your Sunday driving" (Sinnott-Armstrong 2005, p. 312).

Sinnott-Armstrong's empirical assumptions about the individual harmlessness of GHG-emitting activities are open to doubt (see Broome 2019). However, empirical matters aside, his general approach to the abstract issue of collective harm is one which I share. The difference between our

approaches lies not in our conclusions, but in the arguments which we use to arrive at them. The bulk of Sinnott-Armstrong's paper is dedicated to arguing that it is not morally wrong to perform one of many actions which together cause harm, if that action itself is individually harmless. He seems to see no problem in claiming that governments should nonetheless prohibit such actions so as to prevent their harmful joint effects. However, I have pointed out that there are a number of potential problems with this claim. First, if an act is neither singly harmful nor pre-legally wrongful, then this raises questions about the kind of regulation it is permissible to impose up on it. Can legislators justifiably criminalise such an act, or should it be subject instead to civil regulation? And if individuals who perform that act are made liable to bear some cost, then what cost ought that to be, given that their action is individually harmless? Second, certain interpretations of Raz's dependence thesis seem to imply that a political authority cannot justifiably proscribe an act (through either the criminal or the civil law) if, but for that proscription, its members would otherwise have no reason to refrain from it. And finally, if we lack any pre-legal reason to avoid an act, questions arise about whether it would be permissible for lawgivers to disincentivise it through threats of sanction. In the preceding two sections, I have attempted to defuse these difficulties.

Goodin (1995) has also defended a view which bears certain similarities to that presented here. He claims that, sometimes, individuals do not bear responsibility for delivering certain benefits because only a coordinated pattern of actions could deliver them. However, in such cases, Goodin argues that the state bears responsibility for coordinating the required pattern. As such, whilst isolated individuals are not obligated to produce a given benefit, the state has a duty to implement legal regulations which require individuals to act in ways which will jointly produce them.

Like me, Goodin proposes that state officials bear an obligation to introduce legal regulations on certain sorts of behaviour which individuals would otherwise have no reason to avoid. However, his view relies on two assumptions which mine does not. First, he claims that the state's members

share a collective responsibility to deliver benefits which they cannot produce by acting alone. This, for Goodin, is what grounds the state's duty to coordinate collectively beneficial behaviour. Second, he claims that the state constitutes a moral agent in its own right. As such, it takes on the responsibility which its members share collectively (Goodin 1995, pp. 33-37).

The view presented in this chapter relies on neither of these claims. I do not assume that citizens share a collective obligation to avoid combinations of actions which together cause harm; indeed, in the preceding chapter (Section 3.5) I argued that this assumption runs into difficulties. Nor do I rely on the idea that these individuals together comprise a more complex entity, 'the state', which is capable of bearing moral obligations. Rather, my view is consistent with a thoroughgoing individualist ontology of the social world (though it is also consistent with anti-individualist views).[126] Both individualists and anti-individualists can agree that, in societies with an established legal system, there are certain individuals who are endowed with the power to enact, reform, and abolish laws; and they can also agree that these individuals owe it to others to use this power to protect them from at least certain kinds of harm. This, I propose, is what grounds legislators' obligation to regulate collectively harmful behaviour.

5. Objections

I have presented the core of my proposed solution to the paradox of collective harm, and considered some of the ways in which it differs from other similar views. However, I doubt that I have done enough to persuade my reader to endorse this solution. There are a number of possible objections which threaten to render it untenable. I will end this chapter by addressing five potential objections which one may raise against the solution presented in Sections 2 and 3. Each of these worries, I will argue, can be adequately met.

[126] For a helpful discussion of individualism and anti-individualism about social groups, see (Smith 2005).

5.1. Public support

A common objection to views like that which I have defended here is that they overlook the fact that legislative efforts are most effective when they garner support from the general public. Theorists who stress this commonplace of political science tend to think that it counts against purely 'political' or 'institutional' solutions to social problems. What we need, they claim, is not simply a theory which tells politicians what sorts of policy they should be implementing, but also one which shows the general public why they should go along with it. Unless the public understand and support the ideas behind policy, it is unlikely to be successful.[127]

Why might this cause problems for the view I have presented in this chapter? One might formulate the difficulty as follows. My view claims that the legislators ought to regulate acts which individuals would otherwise have no reason to avoid. As such, one might worry that such restrictions would not correspond with people's pre-existing preferences. But if legislation does not overlap with pre-existing preferences, then it is unlikely to be successful.[128] And if it is unlikely to succeed, then why think that implementing restrictions on collectively harmful/individually harmless behaviour would be the best way for legislators to offload their duty of assistance? Thus, in order for my view to succeed, it seems that what we need is an account which shows people why they should want to avoid collectively harmful/individually harmless actions, regardless of whether or not they are illegal. But this is precisely the claim I argue we should reject.

I do not deny that successful legislation requires the support of the public. The preceding paragraph, however, misconstrues that point. It assumes that people have reason to support legislation only if, prior to its implementation, they already have a reason to act as the legislation directs, or incentives, them to act. This is a mistake. I can appreciate that I have reason to support a policy (to vote for it, to campaign for it, not to take steps to block it etc.), if I can see

[127] Iris Marion Young (2011, p. 169) presses a version of this objection against Goodin.
[128] See (Kutz 2000, pp. 181-182).

that legislators have reason to enact it. But part of the point of this chapter has been to demonstrate that a legislator can have a reason to enact laws, even if those subject to them have no pre-legal reason to comply with them.

This is not to deny that it is important to show and convince people that they should support policy which regulates collectively harmful/individually harmless behaviour. In this sense, my account is not purely political; it is a not a 'government house' view which aims to hide the justification of policy from the general public. It is political only in the sense that it holds that people should support the regulations in question not because they already have prior reasons to act as they require, but rather because politicians have reason to enact them.

5.2. Global harm

Another common objection to the sort of view defended here is that it ignores the fact that actions performed under one jurisdiction can collectively cause harm to individuals who reside in a different jurisdiction.[129] For instance, take Sinnott-Armstrong's example of the climate crisis. This is a global crisis; the harm caused by the GHG emissions of one nation will not be confined to its own borders. But if the responsibility to restrict GHG-emitting activities lies primarily with lawgivers, then on what grounds can these individuals restrict behaviour which collectively causes harm to those outside of their jurisdiction?

One way in which to respond to this objection would be to appeal to the role of international law. However, I am sceptical about the prospects of this approach. In order for it to work, there would need to be an identifiable group of international legislators, a compulsory international jurisdiction, and easily imposed sanctions for states which flout international statutes. A system of international law which possessed these features would look less like the system we have now, and more like a kind of 'world order'. But a view which required a world order in order to account for the objectionability of

[129] Young (2011, p. 168) raises this as a further objection to Goodin.

international, collectively caused harm would seem to be unable to explain the wrong of many instances of collective harm which we see now.

There is, however, another, simpler response to this objection. I have argued that legislators are obligated to regulate collectively harmful/individually harmless behaviour because they owe others a duty of assistance. That is, they are obligated to protect others from avoidable harm, whenever feasible. Legislators do not owe this duty exclusively to those within their own jurisdiction. Rather, it is something which everyone owes to one another. As such, it is a mistake to think that lawgivers have no reason to regulate actions which collectively cause harm to people who reside outside of their jurisdiction; this fact does not exempt the lawgivers from their duty to protect these individuals.

One might reply that, assuming that resources are scarce, legislators should prioritise the welfare of those within their own jurisdiction. Given this, legislators may permissibly abstain from assisting members of other jurisdictions, if offering such assistance would require withdrawing assistance from members of their own. My view does not rely on this claim. But even if it is true, it does not entail that legislators are never obligated to regulate behaviour which harms people who reside outside of their own national borders. Such restrictions need not be costly; and if they do not detract from domestic affairs, then they should be implemented.

5.3. Voting

Suppose that two candidates, Good and Bad, are running for presidency. If sufficiently many people vote for Bad, then a great deal of harm will be done: Bad would be a terrible president and would inflict untold damage if elected. By contrast, if enough vote for Good, then this harm will be avoided. Do individual voters have a reason to vote for Good, or at least a reason not to vote for Bad? It might seem intuitive to think that they do. However, this answer is threatened when we consider that the result of an election almost never turns on a single vote; in the vast majority of cases, the result would have been exactly the same had any given individual cast their ballot

differently. But if a single vote is so unlikely to make a difference to the outcome, why think that individuals have a reason not to vote for Bad? Why shouldn't they do so, if it won't make a difference?

The question of how and why we have reason to vote bears certain similarities to the more general problem posed by the paradox of collective harm. As such, some theorists amalgamate them, offering accounts which can give an answer to both. For instance, if correct, Nefsky's (2017) solution to the paradox of collective harm would imply that voters can have a reason to vote for Good. If it is an open possibility that Good will lose the election, then an individual vote for her will play a non-superfluous causal role in averting the harm which Bad's presidency would inflict: even if it later turns out that that vote made no difference to the outcome (see Chapter 4, Section 3.4.2).

My view, however, is different. There seem to be two issues here, which it will be helpful to separate. First, my view offers no answer to the question of how and why we should vote. In fact, it rules out one potential answer to that question. Since I reject the collective harm premise, I deny that individuals have a reason not to vote for Bad which is implied by the fact that that vote will be one of many which will jointly bring about a harmful outcome.

Second, it seems that my account cannot explain the moral objection to the harm which these votes would collectively cause. When applied to the reservoir case, my view implies that it would be morally objectionable for the villagers to be harmed as a joint result of the tippers' actions: either because legislators act wrongfully in failing to prevent those actions, or because the tippers act wrongfully in breaking the law. However, if sufficiently many people voted for Bad, the objectionability of the harm caused by his presidency could not be accounted for in the same way. Presumably, legislators would not be obligated to prohibit people from voting for Bad; to do so would be to flout the conventions of democratic elections. But if so, then we cannot account for the objectionability of the harm in question in terms of legislators' wrongful behaviour. Nor can we account for it in terms

of wrongful disobedience to the law; since any prohibition on voting for Bad would be unjustifiable, it would not be authoritative.[130]

Is any of this a problem? Let's take the second issue first. Is it a problem that my account does not give us a moral objection to the harm which sufficiently many votes for Bad would bring about? Assuming that such harm would be objectionable, this would be problematic only if the best explanation that datum implies the collective harm premise: the harm of Bad's presidency would be objectionable because voters are obligated not to vote for him, in virtue of the harm which those votes will collectively bring about. However, there is an alternative explanation available: the harm inflicted under Bad's presidency would be objectionable not because people ought not to vote for him, but rather because Bad himself ought not to inflict it. Moreover, this explanation seems superior to that which implies the collective harm premise; when elected political leaders act in gratuitously harmful ways, it is they themselves who seem to be the proper objects of criticism, not the people who voted for them. So, the fact that my view does not account for the wrong of harm inflicted by elected officials does not, I suggest, strengthen the abductive case for the collective harm premise. If we endorse my solution to the paradox of collective harm, we do not thereby lose the best account of the wrong of bad government.

Let's return to the first issue. Is it problematic that my view does not imply that voters should not vote for Bad? It would be a problem if endorsing my account meant giving up on the thought that we have any reason at all to cast our votes one way or the other (absent unjustifiable practices like voter bullying). However, it does not imply this. There are answers to the question of how and why we have reason to vote other than those which rely on the collective harm premise.

[130] Why would it be unjustifiable to prohibit people from voting for Bad if he would be such a terrible president? This is a good question. However, a satisfactory answer would take me too far off topic. As such, I will simply assume what seems most problematic for my own account: that it would be wrong for legislators to use the law to influence how individuals cast their votes.

Notice that elections do not have precisely the same structure as cases like that of the reservoir. In Chapter 4 (Sections 2.3 and 3.1) I claimed that the reservoir is an example of a *non-triggering* case: there is no number *n* such that if *n* waste-tippings are performed, then harm suffered by the villagers will be worse than it would have been had *n*-1 waste-tippings occurred. However, elections are not like this; they are *triggering* cases. Consider again the election between Good and Bad. Here, there is some number *n* such that if *n* votes are cast for one of the two candidates, then the result will go one way, whilst had *n*-1 votes been cast for that candidate, the result would have been different. For instance, suppose that the relevant electorate is 10,000,001 in size, that the turn-out rate is 100%, and that a simple first-past-the-post voting system is used. Given this, if 5,000,001 individuals vote for Bad, then the result (Bad wins) will be different to what it would have been had only 5,000,000 individuals voted that way (Good wins).

I also noted in Chapter 4 (Section 3.1) that in triggering cases, individuals can have risk-based reasons to refrain from one of a plurality of collectively harmful actions, even if that action is in fact harmless. Given that no one knows how others will behave, there is a non-zero probability that that action will trigger a greater level of harm. So, if that probability is sufficiently high, or if the harm risked is sufficiently great, then the individuals in question will have a strong reason to refrain from their respective actions.

We can apply this reasoning to the case of Good and Bad as follows. Suppose that no voter possesses reliable information about how others will vote; from their point of view, all vote distributions are equally likely. Given this, the probability that any given vote for Bad will trigger his victory is one in 10,000,001. This is very small. However, the harm which Bad's presidency would inflict is very great, far greater than that which would occur under Good's presidency. As such, the harm risked by a single vote for Bad may well be sufficiently great to give individuals a risk-based reason of not insignificant strength to refrain from voting for him: even if it later turns out that a given vote for Bad would have made no difference to the result.

So, rejecting the collective harm premise does not necessarily imply that voters have no reason to cast their votes one way or another. There are other ways of accounting for the intuition that we can have reasons to use our votes in certain ways which do not appeal to that premise; the risk-based account is a possible example. Nefsky (2017, p. 2748) raises doubts about whether the risk-based account gives us the most plausible picture of why and how we should vote. It does not seem, she says, that my reason for voting for Good is that there is tiny probability that my vote will make a difference to the result. These doubts have most traction, however, only if we can give an alternative account of what that reason might be. Nefsky suggests that her own solution to the paradox of collective harm can give an answer to that question: my reason for voting for Good is that doing so can play a non-superfluous causal role in bringing about her victory. However, I argued in Chapter 4 (Section 3.4.2) that Nefsky's account faces a difficulty. If this is correct, then she will be unable to better the risk-based account of how and why we should vote. Moreover, notice that I am not committed to the claim that the risk-based view offers the best account of this. All I am committed to is the claim that an explanation of our reasons for voting cannot appeal to the collective harm premise. So long as it does not rely on the truth of this premise, any such explanation is admissible.[131]

5.4. Many legislators

The view presented in this chapter accounts for the wrong of collective harm by appealing to the role of legislators. When faced with non-triggering cases, like that of the reservoir, such individuals are in a position to protect others from being harmed: either by issuing morally binding legal directives, or by making justifiable threats and offers. Since we are typically obligated to

[131] Geoffrey Brennan and Loren Lomasky (1989), for instance, argue that voters can have expressive reasons to vote a particular way: by voting for a given candidate, a voter expresses their support for them, even if their ballot makes no difference to the election result. One account of our reasons to vote which my view may rule out is Alvin Goldman's (1999). Goldman argues that even non-swing votes can make a partial causal contribution to the outcome of an election, and that voters whose ballots (or apathy) contribute causally to a bad election result can be subject to moral criticism on that basis. This view seems to rely on the collective harm premise: if a vote for a given candidate is one of several which will together cause a harmful election result, then this provides an individual with a reason not to vote that way.

protect others from avoidable harm where possible, these legislators have a moral obligation to take steps to prevent people from being harmed as a cumulative result of collectively harmful/individually harmless behaviour.

This view might seem to work fine if we make the simplifying assumption that the role of legislator is occupied by just one individual. This person is obligated to exercise their legislative powers in ways which protect people from being harmed, and as such (absent countervailing considerations) ought to implement legal regulations on collectively harmful behaviour. However, political systems are rarely organised in this way. Much more common, at least in democratic societies, are systems in which a plurality of individuals occupy legislative positions. These individuals might vote on legislative proposals, with proposals which receive a majority backing being selected for further rounds of review and voting, and finally being instituted as law only once they have garnered sufficient support at each of these stages.

Such arrangements seem to cause a serious problem for my view. If legislative proposals become law by receiving the support of sufficiently many of a plurality of individuals, then it will not be the case that any one person possesses the power to create new laws. Rather, this will be a capacity possessed jointly by several individuals. But if so, then no one individual will be in a position protect others from the cumulative effects of collectively harmful/individually harmless behaviour through implementing changes to the law. No single legislator will possess the power to issue prohibitions on such behaviour, or to introduce incentives for people to avoid it. All they will be able to do is submit a proposal for such regulations to parliament, and cast their vote in favour of them.

So, in order for my account to deliver the result that collectively caused harm can be morally objectionable under political systems in which legislative proposals are voted on by a plurality of legislators, it needs to be the case that such legislators are (at least sometimes) morally obligated to propose regulations on collectively harmful/individually harmless behaviour, and to vote in favour of such proposals. But what could ground an obligation on the

part of legislators to vote for regulations on collectively harmful behaviour? It can't be that, by voting as such, an individual legislator will protect some other person from avoidable harm. As was pointed out in the preceding subsection, as the size of a voting body increases, the probability that an individual vote will make any difference to the outcome of the ballot approaches zero. Thus, in all likelihood, a given legislator will not protect others from avoidable harm by voting for such a proposal.

At this point, one might worry that the only way in which to fill the lacuna will be to appeal to the collective harm premise: legislators are obligated (at least) not to vote against proposals to regulate collectively harmful behaviour, since many such actions will have the cumulative result that avoidable harm is inflicted upon innocent people; and this in turn implies that we have a reason to refrain from an action if it will be one of several which collectively cause harm to others. If this is correct, then the view presented here collapses. What was initially presented as a rival explanation of the objectionability of collectively caused harm will in fact turn out to rely on the collective harm premise. As such, far from undermining the abductive case for it, the account presented in this chapter would actually *strengthen* the case for that premise.

However, this is to overlook what was established in the preceding subsection. There, I pointed out that ballots are unlike the sorts of case which raise the paradox of collective harm; they are triggering cases. For this reason, individual voters can have risk-based reasons to cast their votes in certain ways; even if a given vote will not in fact make a difference to the outcome, there is a risk that it *might*. This account can be applied to cases in which legislators vote on proposals to regulate collectively harmful behaviour. When the probability that a single vote might make a difference is sufficiently high, or the harm which stands to be caused sufficiently bad, legislators will have a strong risk-based reason, perhaps strong enough to obligate them, to vote in favour of such proposals (or at least not to vote against them).

Thus, the precise mechanics of the view presented in this chapter will depend on the design of the relevant political system. On the simplifying assumption

that there is just one legislator, the objectionability of collectively caused harm is accounted for in terms of that individual's obligation to use their legislative powers to protect others from avoidable harm. However, on the more realistic assumption that the legislature is constituted by a plurality of individuals who vote on legislative proposals, the explanation is slightly different. In a sense, non-triggering cases, like that of the reservoir, are converted into triggering cases: specifically, ballots on whether or not to implement legal regulations on the collectively harmful/individually harmless series of actions which stand to be performed in the relevant non-triggering case. In such ballots, legislators can have risk-based obligations to vote in favour of the regulations in question. Thus, when sufficiently many legislators transgress this obligation that the regulations are not implemented, and harm is caused in the relevant non-triggering case, such harm is morally objectionable: since the legislators were obligated, and yet failed, to take precautions to prevent it.

Notice, however, that this view becomes more unstable as the size of the legislature increases. The more individuals who can vote on legislative proposals, the less likely it will be that any one vote will make a difference to the outcome of the ballot. Thus, whilst members of a smaller legislature might have strong risk-based reasons to vote for regulations on a given series of collectively harmful/individually harmless actions, members of a larger legislature might have comparatively weak reasons to vote for precisely the same regulations: even if the severity of the harm which stands to be caused is held fixed.

There are two ways of looking at this implication. One might take it to be a problem for my account. If harm caused by some collectively harmful/individually harmless series of actions is objectionable, then it shouldn't make a difference how large the legislature of the relevant political system is. According to this objection, the view given here makes the objectionability of collectively caused harm sensitive to the wrong sorts of consideration.

But there is another way of looking at the issue. Instead of treating it an objection to my view, we might instead construe this implication as an objection to certain sorts of institutional design. One of the aims of a political system, we can suppose, is to assign certain responsibilities to identifiable individuals for promoting the legitimate interests of those who live under that system. As such, one of the standards by which we can assess different ways of arranging a political system is how successfully it assigns such responsibilities. More specifically, if a political system fails to assign to some identifiable individual/s the responsibility for protecting certain people's legitimate interests in some circumstance in which they are under threat, then that system will fare worse according to this standard than an alternative system which does assign such responsibilities. In a system with a very large legislature, legislators might not be obligated to vote in favour of certain proposals to regulate collectively harmful/individually harmless behaviour, as they otherwise might be if the legislature was smaller. Thus, under the smaller legislature, responsibilities to take steps to protect certain legitimate interests get assigned to identifiable individuals. Under the larger legislature, by contrast, those same responsibilities do not get assigned. As such, we have (defeasible) grounds for preferring political arrangements with smaller legislatures to those with very large ones: namely, that systems with smaller legislatures can assign responsibilities for protecting individuals from collectively caused harm, whilst systems with larger legislatures may fail to do so in some instances.[132]

Perhaps this response will fail to convince some of my readers. But if so, recall that my view is not wedded to the risk-based account of our reasons to vote in ballots. All I am committed to is the claim that such an account must not rely on the collective harm premise. Thus, if there are alternative ways of making sense of legislators' reasons for voting in favour of certain legislative proposals which do not hang on this premise, then I am happy to accept them.[133]

[132] Thanks to Niko Kolodny for suggesting this way of looking at the issue.
[133] See note 131.

5.5. Non-legal societies

As presented thus far, my account of the wrong of collective harm has relied on the assumption that such harm is caused by the actions of individuals who are subject to some sort of legal system. For instance, when accounting for the objectionability of the harm caused to the villagers by the actions of the tippers, I explicitly assumed that all of these individuals fell under the jurisdiction of a common legal system, and thus that there were some further individuals who possessed the power to legislate against the tippers' actions. However, one might question this assumption. Couldn't it be the case the that the tippers and the villagers live under some pre-, or even post-, legal society?

Indeed, it could. So, let's now assume that the tippers and the villagers inhabit some society in which there is nothing which could conceivably be described as a legal system. Does this assumption create problems for my view? It would not be a problem if it was not intuitive to judge that, under such conditions, harm caused collectively by individually harmless actions is morally objectionable. Perhaps this intuition would not hold under certain pre-legal conditions: a Hobbesian state of nature, for instance. However, it seems not implausible to suppose that there will be at least some possible non-legal conditions under which we will be inclined to judge that such harm is objectionable.

Given this supposition, the following problem arises. My view has it that we should account for the objectionability of any harm caused collectively by the tippers' actions by appealing to the responsibilities of legislators (or the authority of the law). However, in the society under which we are now assuming the tippers and the villagers live, there are neither laws nor legislators. Nevertheless, the harm suffered by the villagers as a result of the tippers' actions is objectionable. How, then, are we to account for this, without appealing to the collective harm premise?

The first thing to note is that there are a number of ways in which individuals can influence how others behave, without exercising legislative powers. For

instance, I might possess a certain stock of goods which I can distribute to others. If these goods a sufficiently desirable, then I will be able to give individuals reasons to act in certain ways by offering to distribute my goods on the condition that they act in those ways. Alternatively, I may be able to give individuals reasons to act as such by withholding my goods. Or again, it might be that I can get a group of people to do certain things simply by telling them to do it, or by doing it myself: that is, they may simply be disposed to do as I do and say (this would be less like a case in which I can give others reasons to do things, and more like one in which I am able to elicit certain non-rational behavioural prompts).

If an individual can prevent others from engaging in collectively harmful behaviour through any of these means, without acting impermissibly, then they may be morally obligated to do so; by taking such steps, they can protect others from avoidable harm. In this way, we can account for the wrong of collective harm without appealing either to the responsibilities of legislators, or to the collective harm premise. So long as there are individuals who are capable of permissibly influencing the tippers' behaviour, the objectionability of the harm caused to the villagers can be explained in terms of the responsibilities which those individuals have to protect others.

At this point one might wonder why I have emphasised the role of the law and legislators. If we can give a rival explanation of the wrong of collective harm without appealing to the responsibilities of legislators, then why appeal to them in the first place? The answer is that the law is a particularly general and robust means of influencing people's behaviour. By enacting or revising certain laws, legislators can create new reasons for action for large numbers of people. Moreover, the class of people to whom a given system of law applies is relatively stable; birth and death are its primary routes of entry and exit. As such, the individuals who have control over a system of law will have primary responsibility for offering protection from others' harmful behaviour; once they have offloaded this responsibility, other individuals who also exercise some behavioural influence may well be absolved of their responsibility. However, when there is no legal system, and thus no

individuals who possess such general and robust influence over people's behaviour, the duty of assistance falls to others who exercise more local influence.

Of course, the law is not the only thing which exerts such a general and robust influence over behaviour. Social customs, norms, traditions, mores: all such informal, unwritten rules of social interaction have a very great effect on how we act. Such things can give us various kinds of reason to behave in certain ways. They can give us prudential reasons to conform if costs (disapproval, humiliation, exclusion) are attached to non-conformity.[134] Moreover, some authors (e.g. Owens 2017; Scheffler 2018) have argued that we can have moral obligations to comply with certain social conventions.

However, recall the point of H. L. A. Hart's which I raised at the end of Chapter 1 (Section 3.3). Unlike in the case of law, there exist no direct means by which customs can be introduced, revised, or repealed. Thus, whilst customs, like laws, can exert general and robust influence over people's behaviour, there is no one who occupies a position comparable to that of a legislator in relation to social customs; no one possesses direct control over these informal rules. The upshot of this is that in societies with a recognisable legal system, there is the opportunity for much greater power to shape others' behaviour than there is under a non-legal social system. Whilst there exist a variety of ways in which individuals can influence one another's behaviour in the absence of law, they cannot do so by means of direct control over a general and robust system of rules.

Given this, my account implies that precisely the same collectively caused harm could potentially be either objectionable or unobjectionable, depending on whether or not the society in which it is enacted possesses a legal system.

[134] As Émile Durkheim writes:

> If I do not conform to ordinary conventions, if in my mode of dress I pay no heed to what is customary in my country and in my social class, the laughter I provoke, the social distance at which I am kept, produce, although in a more mitigated form, the same results as any real penalty (Durkheim 1982, p. 51)

If individuals do not possess very great power to prevent the tippers from dumping their waste into the reservoir, then they may not be obligated to do so. But if no one is obligated to take steps to prevent the tippers' actions, then my view will imply that any harm they jointly cause to the villagers will be unobjectionable.

Is this implication a problem for the view presented here? Again, I think there are two ways of looking at it. On the one hand, one could argue that my account makes the objectionability of collectively caused harm sensitive to the wrong kinds of consideration: namely, whether or not it occurs in a society with a legal system. But, on the other, this implication could also be construed as a defect of non-legal social systems. The institution of law enables particular individuals to wield great power over others, by enabling direct control over a general and robust system of rules. Whilst enabling such power has very great dangers, it also comes with benefits: it extends certain individuals' abilities to protect people from others' harmful behaviour. And with such extensions of ability come extensions of responsibility; under a legal system, responsibilities for protecting people's legitimate interests are assigned to identifiable individuals, which may not otherwise be assigned under a non-legal social system. Such extensions of responsibility are a palpable advantage of the institution of law.[135]

[135] Hart makes a similar point. He argues that it is a defect of non-legal social systems that the rules which govern conduct under them are "static": that is, that they cannot be revised by any direct means. These rules, he claims, need to be adaptable to changing social circumstances. Hart argues that legal systems remedy this defect by introducing what he calls "rules of change": second-order rules which stipulate how an individual may enact revisions to the first-order system of rules. See (Hart 2012, pp. 91-99).

Bibliography

Alexander, Larry (1986). Fair Equality of Opportunity: John Rawls' (Best) Forgotten Principle. *Philosophy Research Archives* 11, 197-208.

Anderson, Elizabeth (1999). What is the Point of Equality? *Ethics* 109 (2), 287-337.

Arneson, Richard (1999). Against Rawlsian Equality of Opportunity. *Philosophical Studies* 93 (1), 77-112.

——— (1989). Equality and Equal Opportunity for Welfare. *Philosophical Studies* 56 (1), 77-93.

Arntzenius, Frank and McCarthy, David (1997). Self Torture and Group Beneficence. *Erkenntnis* 47 (1), 129-144.

Bacharach, Michael (2006). Gold, Natalie and Sugden, Robert (eds.) *Beyond Individual Choice: Teams and Frames in Game Theory*. Princeton NJ.: Princeton University Press.

Barnett, Zach (2018). No Free Lunch: The Significance of Tiny Contributions. *Analysis* 78 (1), 3-13.

Bessone, Magali (2019). Colonial Slave Trade and Slavery and Structural Racial Injustice in France: Using Iris Young's Social Connection Model of Responsibility. *Critical Horizons* 20 (2), 161–177.

Binmore, Ken and Voorhoeve, Alex (2006). Transitivity, the Sorites Paradox, and Similarity-Based Decision-Making. *Erkenntnis* 64 (1), 101-114.

Bobzien, Susanne and Rumfitt, Ian (2020). Intuitionism and the Modal Logic of Vagueness. *Journal of Philosophical Logic* 49 (2), 221-248.

Boolos, George (1984). To Be is to be a Value of a Variable (or to be Some Values of Some Variables). *The Journal of Philosophy* 81 (8), 430-449.

Brennan, Geoffrey and Lomasky, Loren (1989). Large Numbers, Small Costs: The Uneasy Foundation of Democratic Rule. In: Brennan, Geoffrey and Lomasky, Loren (eds.) *Politics and Process: New Essays in Democratic Thought.* Cambridge: Cambridge University Press. 42-59.

Brighouse, Harry and Swift, Adam (2014). *Family Values: The Ethics of Parent-Child Relationships*. Princeton NJ.: Princeton University Press.

——— (2009). Legitimate Parental Partiality. *Philosophy and Public Affairs* 37 (1), 43-80.

——— (2006). Equality, Priority, and Positional Goods. *Ethics* 116 (3), 471-497.

Broome, John (2019). Against Denialism. *The Monist* 102 (1), 110-129.

Budolfson, Mark (2019). The Inefficacy Objection to Consequentialism and the Problem with the Expected Consequences Response. *Philosophical Studies* 176 (7), 1711-1724.

Calvino, Italo (1997). Weaver, William (trans.) *Invisible Cities*. London: Vintage Books. Originally published in 1972 as *Le città invisibile*.

Cohen, G. A. (2008). *Rescuing Justice and Equality*. Cambridge MA.: Harvard University Press.

——— (2000). *If You're an Egalitarian, How Come You're So Rich?* Cambridge MA.: Harvard University Press.

——— (1997). Where the Action Is: On the Site of Distributive Justice. *Philosophy and Public Affairs* 26: 3-30.

——— (1992). Incentives, Inequality, and Community. In: Petersen, G. B. (ed.) *The Tanner Lectures on Human Values*, *Volume Thirteen*. Salt Lake City UT.: University of Utah Press. 262-329.

——— (1989). On the Currency of Egalitarian Justice. *Ethics* 99 (4), 906-944.

Cohen, Joshua (2001). Taking the People as They Are? *Philosophy and Public Affairs* 30 (4), 363-386.

Cooter, Robert and Porat, Ariel (2007). Total Liability for Excessive Harm. *The Journal of Legal Studies* 36 (1), 63-80.

Cullity, Garrett (2000). Pooled Beneficence. In: Almeida, Michael (ed.) *Imperceptible Harms and Benefits*. Dordrecht: Kluwer Academic Publishers. 1-23.

Davidson, Donald (1963). Actions, Reasons, and Causes. *The Journal of Philosophy* 60 (23), 685-700.

Dietz, Alexander (2016). What We Together Ought to Do. *Ethics* 126 (4), 955-982.

Duff, Antony (2002). Crime, Prohibition, and Punishment. *Journal of Applied Philosophy* 19 (2), 97-108.

Durkheim, Émile (1982). Halls, W. D. (trans.) Lukes, Stephen (ed.) *The Rules of Sociological Method*. London: MacMillan. First published in 1895 as *Les règles de la méthode sociologique.*

Dworkin, Ronald (1981). What is Equality? Part 2: Equality of Resources. *Philosophy and Public Affairs* 10 (4), 283-345.

Elster, Jon (1982). Marxism, Functionalism, and Game Theory: The Case for Methodological Individualism. *Theory and Society* 11 (4), 453-482.

Estlund, David (1998). Liberalism, Equality, and Fraternity in Cohen's Critique of Rawls. *The Journal of Political Philosophy* 6 (1), 99-112.

Fanciullo, James (2020). What is the Point of Helping? *Philosophical Studies* 177 (6), 1487-1500.

Feinberg, Joel (1984). *Moral Limits of the Criminal Law Vol. 1: Harm to Others*. Oxford: Oxford University Press.

Fine, Kit (1975). Vagueness, Truth and Logic. *Synthese*, 30 (3/4), 265-300.

Finkelstein, Claire (2003). Is Risk a Harm? *University of Pennsylvania Law Review* 151 (3), 963-1001.

Foot, Philippa (2001). *Natural Goodness*. Oxford: Clarendon.

Gardner, John (2018a). Discrimination: The Good, the Bad, and the Wrongful. *Proceedings of the Aristotelian Society* 118 (1), 55-81.

——— (2018b). *From Personal Life to Private Law*. Oxford: Oxford University Press.

Glover, Jonathan (1975). It Makes no Difference Whether or Not I Do It. *Aristotelian Society Supplementary Volume* 49 (1), 171-209.

Goldblatt, David (2018). Ziebinska-Lewandowska, Karolina (ed.) *Structures of Dominion and Democracy*. Paris: Éditions du Centre Pompidou.

Goldman, Alvin (1999). Why Citizens Should Vote: A Causal Responsibility Approach. *Social Philosophy and Policy* 16 (2), 201-217.

Goodin, Robert E. (1995) *Utilitarianism as a Public Philosophy*. Cambridge: Cambridge University Press.

Graff, Delia (subsequently Fara, Delia Graff) (2000). Shifting Sands: An Interest Relative Theory of Vagueness. *Philosophical Topics* 28 (1), 45-81.

——— (2001). Phenomenal Continua and the Sorites. *Mind* 110 (440), 905-936.

Hart, H. L. A. (2012). *The Concept of Law* 3rd ed. Oxford: Oxford University Press.

Haslanger, Sally (2016). What is (Social) Structural Explanation? *Philosophical Studies* 173 (1), 113-130.

Hausman, Daniel (1995). Rational Choice and Social Theory: A Comment. *The Journal of Philosophy* 92 (2), 96-102.

Hedden, Brian (2020). Consequentialism and Collective Action. *Ethics* 130 (4), 530-554.

Hollis, Martin (1984). Positional Goods. *Royal Institute of Philosophy Supplements* 18, 97-110.

——— (1982). Education as a Positional Good. *Journal of Philosophy of Education* 16 (2), 235-244.

Hossack, Keith (2000). Plurals and Complexes. *The British Journal for the Philosophy of Science* 51 (3), 411-443.

Jackson, Frank (1987). Group Morality. In: Pettit, Philip, Sylvan, Richard, and Norman Jean (eds.) *Metaphysics and Morality: Essays in Honour of J. J. C. Smart*. Oxford: Basil Blackwell. pp. 91-110.

Julius, A. J. (2013). The Possibility of Exchange. *Politics, Philosophy and Economics* 12 (4), 361-374.

——— (2003). Basic Structure and the Value of Equality. *Philosophy and Public Affairs* 31 (4), 321-355.

Kagan, Shelly (2011). Do I Make a Difference? *Philosophy and Public Affairs* 39 (2), 105-141.

Kaufmann, Jean-Claude (1998). Alfrey, Helen (trans.) *Dirty Linen: Couples and Their Laundry*. London: Middlesex University Press. Originally published in 1992 as *La trame conjugal: Analyse du couple par son linge.*

Kolodny, Niko (2017). What Makes Threats Wrong? *Analytic Philosophy* 58 (2), 87-118.

——— (2010). Which Relationships Justify Partiality? The Case of Parents and Children. *Philosophy and Public Affairs* 38 (1), 37-75.

Kutz, Christopher (2000). *Complicity: Ethics and Law for a Collective Age*. Cambridge: Cambridge University Press.

Lévi-Strauss, Claude (1963). Jacobson, Claire and Schoepf, Brooke Grundfest (trans.) *Structural Anthropology.* New York NY.: Basic Books. Originally published in 1958 as *L'anthropologie structurale*.

Little, Daniel (1991). *Varieties of Social Explanation: An Introduction to the Philosophy of Social Science*. Boulder CO.: Westview.

Lu, Catherine (2018). Responsibility, Structural Injustice, and Structural Transformation. *Ethics and Global Politics* 11 (1), 42-57.

——— (2011). Colonialism as Structural Injustice: Historical Responsibility and Contemporary Redress. *The Journal of Political Philosophy* 19 (3), 261-281.

McKay, Thomas (2006). *Plural Predication*. Oxford: Clarendon.

McTernan, Emily (2018). Microaggressions, Equality, and Social Practices. *The Journal of Political Philosophy* 26 (3), 261-281.

Munoz Dardé, Véronique (2002). Family, Choice, and Distributive Justice. In: Archard, David and Macleod, Colin M. (eds.) *The Moral and Political Status of Children.* Oxford: Oxford University Press. 253-268.

——— (1999) Is the Family to be Abolished Then? *Proceedings of the Aristotelian Society* 99 (1), 37-56.

Murphy, Liam (1998). Institutions and the Demands of Justice. *Philosophy and Public Affairs* 27 (4), 251-291.

Nagel, Thomas (1991). *Equality and Partiality*. Oxford: Oxford University Press.

Nefsky, Julia (2019). Collective Harm and the Inefficacy Problem. *Philosophy Compass* 14 (4), 12587-n/a.

——— (2018). Consumer Choice and Collective Impact. In: Barnhill, Anne, Budolfson, Mark, and Doggett, Tyler (eds.) *The Oxford Handbook of Food Ethics*. New York NY.: Oxford University Press. 267-286.

——— (2017). How You Can Help, Without Making a Difference. *Philosophical Studies* 174 (11), 2743-2767.

——— (2015). Fairness, Participation and the Real Problem of Collective Harm. In: Timmons, Mark (ed.) *Oxford Studies in Normative Ethics, Volume 5*. Oxford: Oxford University Press. 245-271.

——— (2011). Consequentialism and the Problem of Collective Harm. *Philosophy and Public Affairs*. 39 (4), 364-395.

Neufeld, Blain (2009). Coercion, the Basic Structure, and the Family. *Journal of Social Philosophy* 40 (1), 37-54.

Okin, Susan Moller (2005). 'Forty Acres and a Mule' for Women: Rawls and Feminism. *Politics, Philosophy and Economics* 4 (2), 233-248.

——— (1989). *Justice, Gender, and the Family*. New York NY.: Basic Books.

Oliver, Alex and Smiley, Timothy (2017). *Plural Logic*. 2nd ed. Oxford: Oxford University Press.

Otsuka, Michael (1991). The Paradox of Group Beneficence. *Philosophy and Public Affairs* 20 (2), 132-149.

Owens, David (2017). Wrong by Convention. *Ethics* 127 (3), 553-575.

Pallikkathayil, Japa (2011). The Possibility of Choice: Three Accounts of the Problem with Coercion. *Philosophers' Imprint* 11 (16), 1-20.

Parfit, Derek (1984). *Reasons and Persons*. Oxford: Oxford University Press.

Pinkert, Felix (2015). What If I Cannot Make a Difference (and Know It). *Ethics* 125 (4), 971-998.

Pogge, Thomas (2000). On the Site of Distributive Justice: Reflections on Cohen and Murphy. *Philosophy and Public Affairs* 29 (2), 137-169.

Rawls, John (1999). *A Theory of Justice* 2nd ed. Cambridge MA.: Harvard University Press.

——— (1993). *Political Liberalism*. New York NY.: Columbia University Press.

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

Rees, J. C. (1960). A Re-Reading of Mill on Liberty. *Political Studies* 8 (2), 113-129.

Sangiovanni, Andrea (2018). Structural Injustice and Individual Responsibility. *Journal of Social Philosophy* 49 (3), 461-483.

Satz, Debra and Ferejohn, John (1994). Rational Choice and Social Theory. *The Journal of Philosophy* 91 (2), 71-87.

Scanlon, T. M. (2018). *Why Does Inequality Matter?* Oxford: Oxford University Press.

——— (2008). *Moral Dimensions: Permissibility, Meaning, and Blame*. Cambridge MA.: Harvard University Press.

——— (1998). *What We Owe to Each Other*. Cambridge MA.: Harvard University Press.

——— (1972). A Theory of Freedom of Expression. *Philosophy and Public Affairs* 1 (2), 204-226.

Scheffler, Samuel (2018). Membership and Political Obligation. *The Journal of Political Philosophy* 26 (1), 3-23.

——— (2006). Is the Basic Structure Basic? In Sypnowich, Christine (ed.) *The Egalitarian Conscience: Essays in Honour of G. A. Cohen*. Oxford: Oxford University Press. 102-129.

——— (2003). What is Egalitarianism? *Philosophy and Public Affairs* 31 (1), 5-39.

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

Sewell, William (1992). A Theory of Social Structure: Duality, Agency, and Transformation. *The American Journal of Sociology* 98 (1), 1-29.

Shelby, Tommie (2016). *Dark Ghettos: Injustice, Dissent, and Reform*. Cambridge MA.: Harvard University Press.

Shiffrin, Seana (2012). Harm and its Moral Significance. *Legal Theory* 18 (3), 357-398.

——— (2010). Incentives, Motives, and Talents. *Philosophy and Public Affairs* 38 (2), 111-142.

——— (2004a). Race, Labour, and the Fair Equality of Opportunity Principle. *Fordham Law Review* 72 (5), 1643-1675.

——— (2004b). Egalitarianism, Choice-Sensitivity, and Accommodation. In: Pettit, Scheffler, Samuel, Philip, Smith, Michael, Wallace, R. Jay (eds.) *Reason and Value: Themes from the Work of Joseph Raz.* Oxford: Oxford University Press. 270-302.

——— (2000). Paternalism, Unconscionability Doctrine, and Accommodation. *Philosophy and Public Affairs* 29 (3), 205-250.

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

Sinnott-Armstrong, Walter (2005). It's Not My Fault: Global Warming and Individual Moral Obligations. In: Sinnott-Armstrong, Walter and Howarth, Richard (eds.) *Perspectives on Climate Change: Science, Economics, Politics, Ethics*. Burlington: Emerald Group Publishing. 293-315.

Smith, Thomas (2011). Playing One's Part. *Review of Philosophy and Psychology* 2 (2), 213-244.

——— (2009). Non-Distributive Blameworthiness. *Proceedings of the Aristotelian Society* 109 (1), 31-60.

——— (2005). What is the Hallé? *Philosophical Papers* 34 (1), 75-109.

Spiekermann, Kai (2014). Small Impacts and Imperceptible Effects: Causing Harm with Others. *Midwest Studies in Philosophy* 38 (1), 75-90.

Sunstein, Cass and Thaler, Richard (2008). *Nudge: Improving Decisions About Health, Wealth, and Happiness*. New Haven Conn.: Yale University Press.

Swift, Adam (2003). *How Not to be a Hypocrite: School Choice for the Morally Perplexed Parent*. London: Routledge.

Tadros, Victor (2010). Criminalization and Regulation. In: Duff, R. A., Farmer, Lindsay, Marshall, S. E., Renzo, Massimo, and Tadros, Victor (eds.) *The Boundaries of the Criminal Law*. Oxford: Oxford University Press.163-190.

Tan, Kok-Chor (2004). Justice and Personal Pursuits. *The Journal of Philosophy* 101 (7), 331-362.

Waldron, Jeremy (1987). Mill and the Value of Moral Distress. *Political Studies*. 35 (3), 410-423.

White, Stephen (2017). On the Moral Objection to Coercion. *Philosophy and Public Affairs* 45 (3), 199-231.

Wiggins, David (2006). *Ethics: Twelve Lectures in the Philosophy of Morality*. London: Penguin.

Williams, Andrew (1998). Incentives, Inequality, and Publicity. *Philosophy and Public Affairs* 27 (3), 225-247.

Williams, Bernard (1962). The Idea of Equality. In: Laslett, Peter and Runciman, W. G. (eds.) *Philosophy, Politics and Society* 2nd series. Oxford: Basil Blackwell. 110-131.

Williamson, Timothy (1994). *Vagueness*. London: Routledge.

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

——— (2006a). Responsibility and Global Justice: A Social Connection Model. *Social Philosophy and Policy* 23 (1), 102-130.

——— (2006b). Katrina: Too Much Blame, Not Enough Responsibility. *Dissent* 53 (1), 41-46.

——— (2004). Responsibility and Global Labor Justice. *The Journal of Political Philosophy* 12 (4), 365-388.

——— (2001). Equality of Whom? Social Groups and Judgements of Injustice. *The Journal of Political Philosophy* 9 (1), 1-18.